

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2004, 18:27:09 ; Search time 46 Seconds

(without alignments)

4492.705 Million cell updates/sec

Title: US-09-961-086-1

Perfect score: 3352

Sequence: 1 MSSNNVEPIPVQSGNTNGF.....MIVIFLTAYLKILFLKKYS 655

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTRMBL25:  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phase.\*  
10: sp\_plant.\*  
11: sp\_protein.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_rvdrus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3346	99.8	655	Q96TA8	Q96TA8 homo sapien
2	3342	99.7	655	Q8IX16	Q8IX16 homo sapien
3	3338	99.6	655	Q96LD6	Q96LD6 homo sapien
4	2849.5	85.0	656	Q8MIB3	Q8MIB3 sus scrofa
5	2762	82.4	657	Q7TMS5	Q7TMS5 mus musculu
6	2761	82.4	541	Q86V64	Q86V64 homo sapien
7	2757	82.2	657	Q9R004	Q9R004 mus musculu
8	2754	82.2	657	Q8OW57	Q8OW57 rattus norv
9	2752	82.1	657	Q8OST1	Q8OST1 rattus norv
10	2742	81.8	657	Q8OXF3	Q8OXF3 rattus norv
11	1787.5	53.3	650	Q8KIS	Q8KIS mus musculu
12	847.5	25.3	801	Q8T9L	Q8T9L dictyosteli
13	809	24.1	751	Q93YS4	Q93YS4 arabadopsis
14	808	24.1	687	Q9NH94	Q9NH94 bombyx mori
15	806	24.0	687	Q94960	Q94960 drosophila
16	800.5	23.9	737	Q9FT51	Q9FT51 arabadopsis

17	793	23.7	679	5	Q9BH97	Q9BH97 ceratitis c
18	790	23.6	692	5	P91892	P91892 aedes aegypt
19	783.5	23.4	648	10	Q9C6W5	Q9C6W5 arabadopsis
20	777.5	23.2	679	5	Q8IS30	Q8IS30 bactroceira
21	776.5	23.2	567	10	Q9FG17	Q9FG17 arabadopsis
22	774	23.1	646	10	Q9C6R7	Q9C6R7 arabadopsis
23	772	23.1	670	5	O77423	O77423 bactroceira
24	767.5	22.9	687	10	Q9C8K2	Q9C8K2 arabadopsis
25	766.5	22.9	695	10	Q8LMQ5	Q8LMQ5 oryza sativ
26	763.5	22.8	798	5	Q8T689	Q8T689 dictyosteli
27	759	22.6	703	10	Q8RXN0	Q8RXN0 arabadopsis
28	755	22.5	672	10	Q9LI82	Q9LI82 arabadopsis
29	751	22.4	669	5	Q8WRE2	Q8WRE2 tribolium c
30	749.5	22.4	691	10	Q8RWI9	Q8RWI9 arabadopsis
31	746	22.3	649	10	Q9SIT6	Q9SIT6 arabadopsis
32	745	22.2	725	10	Q9M3D6	Q9M3D6 arabadopsis
33	744	22.2	669	5	Q8WR1	Q8WR1 tribolium c
34	743.5	22.2	1528	5	Q8T677	Q8T677 dictyosteli
35	743	22.2	1328	5	Q9NGP5	Q9NGP5 dictyosteli
36	741.5	22.1	739	10	Q9LFG8	Q9LFG8 arabadopsis
37	739	22.0	678	10	Q9C8J8	Q9C8J8 arabadopsis
38	739	22.0	678	10	Q8GX48	Q8GX48 arabadopsis
39	730.5	21.8	662	10	Q949Y4	Q949Y4 arabadopsis
40	730.5	21.8	662	10	Q84YH5	Q84YH5 arabadopsis
41	730.5	21.8	1520	5	Q8T687	Q8T687 dictyosteli
42	730	21.8	626	5	Q8T684	Q8T684 dictyosteli
43	730	21.8	626	5	Q86HQ2	Q86HQ2 dictyosteli
44	728.5	21.7	692	10	Q7XUM2	Q7XUM2 oryza sativ
45	727.5	21.7	668	10	Q9ARU4	Q9ARU4 oryza sativ

## ALIGNMENTS

### RESULT 1

Q96TA8	PRELIMINARY;	PRT;	655	AA.
ID	Q96TA8			
AC	Q96TA8;			
DT	01-DEC-2001 (TREMBLrel. 19, Created)			
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)			
DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)			
DE	ATP-binding cassette superfamily G (White) member 2 (Hypothetical protein).			
GN	ABC2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE-21201983; PubMed-11306452;			
RA	Komatsu H., Kotani H., Hara Y., Nakagawa R., Matsumoto M., Arakawa H., Nishimura S.;			
RA	"Identification of breast cancer resistant protein/mitoxantrone resistance/placenta-specific, ATP-binding cassette transporter as a transporter of NB-506 and J-107088, topoisomerase I inhibitors with an indolocarbazole structure."			
RL	Cancer Res. 61:2827-2832(2001).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Famreatic carcinoma;			
RA	Strausberg R.;			
RA	Submitted (JAN-2002) to the EMBL/GenBank/DBSJ databases.			
DR	ENBL; ABC051855; BAB46933.1; -			
DR	ENBL; BC021281; AAH21281.1; -			
DR	GO; GO:0016020; C:membrane; IEA.			
DR	GO; GO:0005524; F:ATP binding; IEA.			
DR	GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . . ; IEA.			
DR	GO; GO:000810; P:transport; IEA.			
DR	InterPro; IPR003439; ABC_transporter.			
DR	InterPro; IPR006162; Ppantne S.			
DR	Pfam; PF00005; ABC_tran; 1.			
DR	ProDom; PD000006; ABC_transporter; 1.			



OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Schuetz J.D., Wall A.M., Sampath J., Sorrentino B., Du G.;  
 RT "The Human ABC Transporter, ABCG2, Transports Hoechst 33342 and  
 RL Requires an Intact Walker A Motif";  
 DR Submitt (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 DR ENBL, AY017168; AAG52982.1;  
 DR GO:0016020; C:membrane; IEA.  
 DR GO:0005224; F:ATP binding; IEA.  
 DR GO:0004009; F:ATP-binding cassette (ABC) transporter acti. .; IEA.  
 DR GO:0000166; F:nucleotide binding; IEA.  
 DR GO:0006810; P:transport; IEA.  
 DR InterPro: IPR003593; AAA\_ATPase.  
 DR InterPro: IPR003439; ABC\_transporter.  
 DR InterPro: IPR006162; Pfamne\_S.  
 DR Pfam: PF00005; ABC\_tran; 1.  
 DR ProDom: PD000006; ABC\_transporter; 1.  
 DR PROSITE: PS00012; PHOSPHOPANTETHEINE; 1.  
 DR ATP-binding.  
 SQ SEQUENCE 555 AA; 72288 MW; B3B5DC02C095C4A8 CRC64;  
 Query Match 99.6%; Score 3338; DB 4; Length 555;  
 Best Local Similarity 99.7%; Pred. No. 5.1e-225;  
 Matches 653; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 MSSSNEVEFIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICVRVVKSGFLPCRKPE 60  
 DB 1 MSSSNEVEFIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICVRVVKSGFLPCRKPE 60  
 QY 61 KEILSNINGIMKPGNALILGPTGGKSSLLDLAARKDPSGLSGDVLINGAPRANPKCN 120  
 DB 61 KEILSNINGIMKPGNALILGPTGGKSSLLDLAARKDPSGLSGDVLINGAPRANPKCN 120  
 QY 121 SGVVQDDVVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIQELGLDKVADSKVGT 180  
 DB 121 SGVVQDDVVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIQELGLDKVADSKVGT 180  
 QY 181 QFIRGVSGGERKRTSGMELITDPSILFLDEPTTGLDSTANAVLLLLKMSKQGRITIF 240  
 DB 181 QFIRGVSGGERKRTSGMELITDPSILFLDEPTTGLDSTANAVLLLLKMSKQGRITIF 240  
 QY 241 SIHOPRYSIFKLFDLSLTLASGLMFGHPAQALGYFESAGYHCEAYNNPADFFLDIING 300  
 DB 241 SIHOPRYSIFKLFDLSLTLASGLMFGHPAQALGYFESAGYHCEAYNNPADFFLDIING 300  
 QY 301 DSTAVALNRREDFKATEIIPSKODKPLIEKLAETIVNSFFYKETAELHOLSGGEKKK 360  
 DB 301 DSTAVALNRREDFKATEIIPSKODKPLIEKLAETIVNSFFYKETAELHOLSGGEKKK 360  
 QY 361 ITVPKEISYTTSFCHOLRWWSKRSFKNLGNPQASIAQIIVTVVLGIVGAIYFGLKND 420  
 DB 361 ITVPKEISYTTSFCHOLRWWSKRSFKNLGNPQASIAQIIVTVVLGIVGAIYFGLKND 420  
 QY 421 TGIONRAGVLFFLITNQCRSSVSAVELFVVEKKLFIEHISGYRVSSVFLGLKSLDLLP 480  
 DB 421 TGIONRAGVLFFLITNQCRSSVSAVELFVVEKKLFIEHISGYRVSSVFLGLKSLDLLP 480  
 QY 481 MTMLPSIIFTCIVYFMGLKPKADAFVVMFTLMWVAYSSASSNALAIAAGGSVSVATLL 540  
 DB 481 MRMLPSIIFTCIVYFMGLKPKADAFVVMFTLMWVAYSSASSNALAIAAGGSVSVATLL 540  
 QY 541 MTICFVPMWIFSGLLVNLITIASLWQLQYFSIPRYGFTALQHNFLQNGFCPLGNATGN 600  
 DB 541 MTICFVPMWIFSGLLVNLITIASLWQLQYFSIPRYGFTALQHNFLQNGFCPLGNATGN 600  
 QY 601 NPCNVATCTGEBVLVKQIDLSFWGLWKNHVALACMIVFLTIAYLKLFLKKYS 655

DB 601 NPCNVATCTGEBVLVKQIDLSFWGLWKNHVALACMIVFLTIAYLKLFLKKYS 655  
 RESULT 4  
 QSMIB3 PRELIMINARY; PRT; 656 AA.  
 ID QSMIB3;  
 AC 01-OCT-2002 (TREMBlrel. 22, Created)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Brain multidrug resistance protein.  
 GN BMDP.  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 OX NCBI\_TaxID=9823;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=22050127; PubMed=12054514;  
 RX Eisenblaeetter T., Galla H.J.;  
 RT "A new multidrug resistance protein at the blood-brain barrier";  
 RL Biochem. Biophys. Res. Commun. 293:1273-1278(2002).  
 DR EMBL, AJ420927; CAD12785.1; -.  
 DR FIR, JC7860; JC7860.  
 DR GO:0016020; C:membrane; IEA.  
 DR GO:0005524; F:ATP binding; IEA.  
 DR GO:0004009; F:ATP-binding cassette (ABC) transporter acti. .; IEA.  
 DR GO:0000166; F:nucleotide binding; IEA.  
 DR GO:0006810; P:transport; IEA.  
 DR InterPro: IPR003593; AAA\_ATPase.  
 DR InterPro: IPR003439; ABC\_transporter.  
 DR Pfam: PF00005; ABC\_tran; 1.  
 DR ProDom: PD000006; ABC\_transporter; 1.  
 DR SMART: SM00382; AAA; 1.  
 DR PROSITE: PS00893; ABC\_TRANSPORTER\_2; 1.  
 DR PROSITE: PS00012; PHOSPHOPANTETHEINE; 1.  
 DR ATP-binding.  
 KW SEQUENCE 656 AA; 72392 MW; 118ADD5B53D9D67F CRC64;  
 Query Match 85.0%; Score 2849.5; DB 6; Length 656;  
 Best Local Similarity 84.3%; Pred. No. 4e-189;  
 Matches 553; Conservative 44; Mismatches 58; Indels 1; Gaps 1;  
 QY 1 MSSSNEVEFIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICVRVVKSGFLPCRKPE 60  
 DB 1 MSSSNEVEFIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICVRVVKSGFLPCRKPE 60  
 QY 61 KEILSNINGIMKPGNALILGPTGGKSSLLDLAARKDPSGLSGDVLINGAPRANPKCN 120  
 DB 61 KEILSNINGIMKPGNALILGPTGGKSSLLDLAARKDPSGLSGDVLINGAPRANPKCN 120  
 QY 121 SGVVQDDVVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIQELGLDKVADSKVGT 180  
 DB 121 SGVVQDDVVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIQELGLDKVADSKVGT 180  
 QY 181 QFIRGVSGGERKRTSGMELITDPSILFLDEPTTGLDSTANAVLLLLKMSKQGRITIF 240  
 DB 181 QFIRGVSGGERKRTSGMELITDPSILFLDEPTTGLDSTANAVLLLLKMSKQGRITIF 240  
 QY 241 SIHOPRYSIFKLFDLSLTLASGLMFGHPAQALGYFESAGYHCEAYNNPADFFLDIING 300  
 DB 241 SIHOPRYSIFKLFDLSLTLASGLMFGHPAQALGYFESAGYHCEAYNNPADFFLDIING 300  
 QY 301 DSTAVALNRREDFKATEIIPSKODKPLIEKLAETIVNSFFYKETAELHOLSGGEKKK 359  
 DB 301 DSAAVLSRADDEGAQEPPEPEKOTFLIDKLAFTYNSFFKDTKVELDQSGGGRKK 360  
 QY 360 KITVPKEISYTTSFCHOLRWWSKRSFKNLGNPQASIAQIIVTVVLGIVGAIYFGLKND 419  
 DB 361 KSSVYKEVYTTTSFCHOLRWWSKRSFKNLGNPQASIAQIIVTVVLGIVGAIYFGLKND 420  
 QY 420 STGIQNRAGVLFFLITNQCRSSVSAVELFVVEKKLFIEHISGYRVSSVFLGLKSLDLL 479

```

421  PSQIQNRAGVLFELITNQCFSSVSAVELLWVEKLFIEHYISGYRVSFFGKLLSDLL 480
480  EMTLPSIIFTCIVFPMGLGPKADAFVMMFTLMMVAYSASSMALAIAAGOSVVSATL 539
481  EMTLPSIIFTCIVFPMGLGPKADAFVMMFTLMMVAYSASSMALAIAAGOSVVSATL 540
540  LMTICFVFMIFSGLLVNLITIASMSLWQYSPRIYFTALQHNFLGQFCPLNATG 599
541  LMTISFVFMIFSGLLVNLITIASMSLWQYSPRIYFTALQHNFLGQFCPLNATG 600
600  NNPVATCTGEEYLVKQIDLSPLMGLWKNHVALACMIIVITLTIAYLKLFLKKYS 655
601  NNTCSFAICTGAEYLENQGSLISAWGLWQNHVALACMIIVITLTIAYLKLFLKKYS 656

RESULT 5
Q7TMS5
ID Q7TMS5 PRELIMINARY; PRT; 657 AA.
AC Q7TMS5;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE ATP-binding cassette, sub-family G, member 2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6Ncr; TISSUE=Hematopoietic Stem Cell;
RX MEDLINE=22398257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heien F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Boeak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RL and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6Ncr; TISSUE=Hematopoietic Stem Cell;
RX Strausberg R.;
RA Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC053730; ARL53730.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. .; IEA.
DR GO; GO:0006810; F:transport; IEA.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR006162; Ppantre_S.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PDC00006; ABC transporter; 1.
DR PROSITE; PS00012; PROSOPANTHETINE; 1.
KW ATP-binding.
KW NON_TER
FT SEQUENCE 541 AA; 72977 MW; DCD70C5D9FA2BA5F CRC64;
SQ SEQUENCE 657 AA; 72977 MW; DCD70C5D9FA2BA5F CRC64;

Query Match 82.4%; Score 2762; DB 11; Length 657;
Best Local Similarity 81.5%; Pred. No. 4.7e-183;
Matches 536; Conservative 52; Mismatches 56; Indels 4; Gaps 3;

QY 1 MSSNVEVFPVSGNGTNGPPATASNDLKAFTEGAVLSFNICVYKVLKSGFLPCRPVYE 60
DB 1 MSSNDHVLVPMSONRNLPRNTNSRAVRTLAEGDVLSPFHITVYKVKSGFL-VKRTVE 59
QY 61 KEILSNINGIMKGLNAILGPTGGKSSLDVLAARKDPSGLSGDVLINGAPRANFKCN 120
DB 60 KEILSDINGIMKGLNAILGPTGGKSSLDVLAARKDPSGLSGDVLINGAPRANFKCC 119

```

```

QY 121 SGTVVODVVMGTLTVRENLQPSAALRLATTTNHEKNERINRVIOELGLDKVADSKVCT 180
DB 120 SGTVVODVVMGTLTVRENLQPSAALRLPTTWNHEKNERINTIIRKELGLEKADSKVCT 179
QY 181 QFIRGVSGGERKRTSICMELITDPSILFDEPTTGLDSDSTANAVLLLLKRMKQGRITIF 240
DB 180 QFIRGISGGERKRTSICMELITDPSILFDEPTTGLDSDSTANAVLLLLKRMKQGRITIF 239
QY 241 SIHQPRYSIFKLPSDITLTLASGLMHFGPAQALGYFESAGYHCEAYNNPADFFLDLNG 300
DB 240 SIHQPRYSIFKLPSDITLTLASGLVPHGPAQALEYFASAGYHCEAYNNPADFFLDVING 299
QY 301 DSTAVALNREE-DFKATEIIEPSKQDKPLIEKLAIEYVNSFFYKTKAELHQLSGGEKKK 359
DB 300 DSAVNLNREEQNEANKTEEPSKGEKPVLENSEYVNSIAYGETKAELEDQPGAQEKK 359
QY 360 KITVFKESYTTSCFCHQLRWKRSFKNLGNPQASIAQIIVTVVLGLVIGALYFGLKND 419
DB 360 GTSAFKEPVYVTSFCHQLRWIARRSPKLLGNPQASVAQIIVTVVLGLVIGALYFGLKYD 419
QY 420 STGIONRAGVLFELITNQCFSSVSAVELFVVEKLFIEHYISGYRVSFFGKLLSDLL 479
DB 420 AAGQNRAGVLFELITNQCFSSVSAVELFVVEKLFIEHYISGYRVSFFGKLLSDLL 479
QY 480 PMTMLPSIIFTCIVFPMGLGPKADAFVMMFTLMMVAYSASSMALAIAAGOSVVSATL 539
DB 480 PMRPLPSVIPTCVLYFPMGLGPKADAFVMMFTLMMVAYSASSMALAIAAGOSVVSATL 539
QY 540 LMTICFVFMIFSGLLVNLITIASMSLWQYSPRIYFTALQHNFLGQFCPLNATG 599
DB 540 LMTIAFVFMIFSGLLVNLITIASMSLWQYSPRIYFTALQHNFLGQFCPLNATG 599
QY 600 NNPVATCTGEEYLVKQIDLSPLMGLWKNHVALACMIIVITLTIAYLKLFLKKYS 655
DB 600 NSTCVNSVAICTGNEYLINGQIELSPMGLWKNHVALACMIIVITLTIAYLKLFLKKYS 657

RESULT 6
Q86V64
ID Q86V64 PRELIMINARY; PRT; 541 AA.
AC Q86V64;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE ATP-binding cassette sub-family G member 2 (Fragment).
GN ABCG2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Zhang W., Petrovic J.-M., Andrade M.F., Zhang H., Stanimirovic D.B.;
RT "The Expression and Functional Characterization of ABCG2 in Brain
RT Endothelial Cells and Vessels.";
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY288307; AAP31310.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. .; IEA.
DR GO; GO:0006810; F:transport; IEA.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR006162; Ppantre_S.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PDC00006; ABC transporter; 1.
DR PROSITE; PS00012; PROSOPANTHETINE; 1.
KW ATP-binding.
KW NON_TER
FT SEQUENCE 541 AA; 60450 MW; 050B2742C6A3F66C CRC64;
SQ SEQUENCE 541 AA; 60450 MW; 050B2742C6A3F66C CRC64;

Query Match 82.4%; Score 2761; DB 4; Length 541;
Best Local Similarity 99.8%; Pred. No. 4.2e-183;
Matches 540; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 115 ANFKNSGYVQDDVVMGTLTVRENLQFSAALRLATTMTNHEKNERINRVIOELGLDKA 174
DB 1 ANFKNSGYVQDDVVMGTLTVRENLQFSAALRLATTMTNHEKNERINRVIOELGLDKA 60
QY 175 DSKVGTQFIRGSGGKRTSISGMELITDPSILFDEPTTGDDSTANAVILLKRMKQ 234
DB 61 DSKVGTQFIRGSGGKRTSISGMELITDPSILFDEPTTGDDSTANAVILLKRMKQ 120
QY 235 GRTIISIHOPRYSIKLFDLSLTLASGRIMFHPGPAQALGYFESAGYHCEAYNPPDF 294
DB 121 GRTIISIHOPRYSIKLFDLSLTLASGRIMFHPGPAQALGYFESAGYHCEAYNPPDF 180
QY 295 LDIINGDSTAVLNREEDFKATEIIEPSKQDKPLIEKLAIEIYVNSSFYKETAELHQLSG 354
DB 181 LDIINGDSTAVLNREEDFKATEIIEPSKQDKPLIEKLAIEIYVNSSFYKETAELHQLSG 240
QY 355 GEKKKKTIVPEKISYTSFCHQURWVSKFSKNLGNPOASTAQIIVTVGLVIGAIYF 414
DB 241 GEKKKKTIVPEKISYTSFCHQURWVSKFSKNLGNPOASTAQIIVTVGLVIGAIYF 300
QY 415 GLKNDSTGIQNRAGVLFLLTNNOCFSSVSAVELFVVEKKLFTHIYISGYRVSSYFLGKL 474
DB 301 GLKNDSTGIQNRAGVLFLLTNNOCFSSVSAVELFVVEKKLFTHIYISGYRVSSYFLGKL 360
QY 475 LSDLLPMTLPSIIFTCIVVFMGLKPKADAFVMMFTLMWYASASSMALAIAGQSVV 534
DB 361 LSDLLPMTLPSIIFTCIVVFMGLKPKADAFVMMFTLMWYASASSMALAIAGQSVV 420
QY 535 SVATLLMTICFVFMFIFSGLLVNLTTIASWLSWLYFSIPRYGPTALQHNFPLOGQFCG 594
DB 421 SVATLLMTICFVFMFIFSGLLVNLTTIASWLSWLYFSIPRYGPTALQHNFPLOGQFCG 480
QY 595 LNATGNPNVYACTGEYLVKQIDLSPWGLWKNHVALACMIVIFLTAYLKLFLKKY 654
DB 481 LNATGNPNVYACTGEYLVKQIDLSPWGLWKNHVALACMIVIFLTAYLKLFLKKY 540
QY 655 S 655
DB 541 S 541

RESULT 7
Q9R004 Q9R004 PRELIMINARY; PRT; 657 AA.
AC Q9R004;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-OCT-2003 (T-EMBlrel. 25, Last annotation update)
DE Breast cancer resistance protein 1.
GN ABCG2 OR BCRP1
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10080;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB; TISSUE=Liver;
RX MEDLINE=99413474; PubMed=10485464;
RA Allen J.D., Brinkhuis R.F., Wijnholds J., Schinkel A.H.;
RT "The mouse Bcrp/Mxr/Abcg Gene: amplification and overexpression in
RT cell lines selected for resistance to topotecan, mitoxantrone, or
RT doxorubicin."
RL Cancer Res. 59:4237-4241(1999).
DR EMBL; AF140218; AAD54216.1; -.
DR MGD; MGI:1347061; Abcg2.
DR GO; GO:0016020; C.membrane; IEA.
DR GO; GO:0005524; P.ATP binding; IEA.
DR GO; GO:0004009; F.ATP-binding cassette (ABC) transporter acti. .; IEA.
DR GO; GO:0006166; F.nucleotide binding; IEA.
DR GO; GO:0006810; P.transport; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.

```

```

InterPro; IPR006162; Ppantne_S.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD00006; ABC transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00893; ABC TRANSPORTER 2; 1.
DR PROSITE; PS00912; PHOSPHOPANTHEINE; 1.
KW ATP-binding.
SQ SEQUENCE 657 AA; 73021 MW; 207B70BC272CC05 CRC64;

Query Match 82.2%; Score 2757; DB 11; Length 657;
Best Local Similarity 81.5%; Pred. No. 1e-182;
Matches 536; Conservative 51; Mismatches 67; Indels 4; Gaps 3;

QY 1 MSSNNVEFIPVPSQGTNGFPATASNDLKAFTEGAVLSFHNICRYRVLKSGFLPCRPVE 60
DB 1 MSSNDHVLVPMSQRNNGLPRMNSRAVRTLAEQDVLSEFHIIVRVKYSGL-VRKIVE 59
QY 61 KEILSNINIMKPGNALILGPTGGGKSLLDVLAARKDPSGLSDVLINGAPRANPKCN 120
DB 60 KEILSDINIMKPGNALILGPTGGGKSLLDVLAARKDPSGLSDVLINGAPQPAHFCC 119
QY 121 SGVYQDDVVMGTLTVRENLQFSAALRLATTMTNHEKNERINRVIOELGLDKVADSKVGT 180
DB 120 SGVYQDDVVMGTLTVRENLQFSAALRLATTMTNHEKNERINTIIEKLEKADSKVGT 179
QY 181 QTRGVSGGERKRTSIGMELITDPSILFDEPTTGDDSTANAVILLKRMKQRTIIF 240
DB 180 QTRGVSGGERKRTSIGMELITDPSILFDEPTTGDDSTANAVILLKRMKQRTIIF 239
QY 241 SIHQPRYSIFKLFDLSLTLASGRIMFHPGPAQALGYFESAGYHCEAYNPPDFLDIING 300
DB 240 SIHQPRYSIFKLFDLSLTLASGRIMFHPGPAQALGYFESAGYHCEAYNPPDFLDVING 299
QY 301 DSTAVANREB-DFKATIIIPSKODKPLIEKLAIEIYVNSSFYKETAELHOLSGGKK 359
DB 300 DSAVNLNREEDQNEANKTEPSKGEKPEIENLSEFYINSIAYIETKAEQLDLPAGEKK 359
QY 360 KITVFKESITTSFCHQURWVSKFSKNLGNPOASTAQIIVTVGLVIGAIYFGLKND 419
DB 360 GT-SAPKEPVVTSFCHQURWVSKFSKNLGNPOASTAQIIVTVGLVIGAIYFGLKYD 419
QY 420 STGIONRAGVLFLLTNNOCFSSVSAVELFVVEKKLFTHIYISGYRVSSYFLGKLSDL 479
DB 420 AAGMQRAGVLFLLTNNOCFSSVSAVELFVVEKKLFTHIYISGYRVSSYFLGKMSDLL 479
QY 480 PMTMLPSIIFTCIVVFMGLKPKADAFVMMFTLMWYASASSMALAIAGQSVSVATL 539
DB 480 PMRFLPSVIFTCILYFMLGLKKTVDAFFIMFTLMWYATASSMALAIATGQSVSVATL 539
QY 540 LMTICFVFMFIFSGLLVNLTTIASWLSWLYFSIPRYGPTALQHNFPLOGQFCPLNATG 599
DB 540 LMTIAFVFMFIFSGLLVNLRTIGPWSLWQYFSIPRYGPTALQHNFPLOGQFCGFNWD 599
QY 600 NNPC--NYATCTGEEYLVKQIDLSPWGLWKNHVALACMIVIFLTAYLKLFLKKYS 655
DB 600 NSTCVNSVAICTGNEYLNQIELSPWGLWKNHVALACMIIIFLTAYLKLFLKKYS 657

RESULT 8
Q9R0W57 Q9R0W57 PRELIMINARY; PRT; 657 AA.
AC Q9R0W57;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE ABC transporter ABCG2.
GN ABCG2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.

```

[illegible]

Db	420	PTGMQNRAGVFFLLT	TTNQCTSVSAVEL	FVVEKLTPIHE	YISGIVRVS	SSFFGKGLVSULL	47
Qy	480	PMTMLPSIIPTCI	VYFVFMGLGKPKAD	FFVMFTLMWVAYS	SASNMA	LAIAAGOSVSVATL	539

RESULT 9	
Q80ST1	
ID	Q80ST1
AC	Q80ST1

```

DT 01-JUN-2003 (trEMBLrel. 24, Created)
DT 01-JUN-2003 (trEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (trEMBLrel. 25, Last annotation update)
DE ATP-binding cassette protein G2 transcript variant B (ATP-binding
DE cassette protein G2 transcript variant C) (ATP-binding cassette
DE protein G2 transcript variant A).
GN ABCG2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBL_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=liver;
RA Yabuuchi H., Ishikawa T.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY089996; AA09106.1; -
DR EMBL; AY089997; AA09107.1; -
DR EMBL; AY089998; AA09108.1; -
DR GO: GO:0016020; C:membrane; IEA.
DR GO: GO:0005224; F:ATP binding; IEA.
DR GO: GO:0004009; F:ATP-binding cassette (ABC) transporter acti. .; IEA.
DR GO: GO:0003166; F:nucleotide binding; IEA.
DR GO: GO:0006810; P:transport; IEA.
DR InterPro: IPR003593; AAA ATPase.
DR InterPro: IPR003439; ABC transporter.
DR InterPro: IPR006162; Ppantne_S.
DR Pfam: PF00005; ABC_tran; 1.
DR ProDom: PD000006; ABC_transporter; 1.
DR SMART: SM00382; AAA; 1.
DR PROSITE: PS0893; ABC_TRANSPORTER_2; 1.
DR PROSITE: PS00012; PHOSPHOPANTETHEINE; 1.
KW ATP-binding.
KW ATP-binding.
SQ
SEQUENCE 657 AA; 72960 MW; E194871B1C1AC201 CRC64;

Query Match 82.1%; Score 2752; DB 11; Length 657;
Best Local Similarity 80.9%; Pred No. 2.3e-182;
Matches 532; Conservative 53; Mismatches 69; Indels 4; Gaps 3;

QY 1 MSSNVVEFIPVSGQTNGFPATASNDLKAFTEGAVLSFHNI CYRVKXKSGFLPCRPVE 60
DB 1 MSSNDHVLVPMQSRKNKGLPGMSSRGARTLAEGDVLSPHHITYRVKXKSGFL-VRKTAE 59
QY 61 KEILSNITNGIMKPGNALILGPTGGKSSLLDLVLAARKDPGSLGSDVLINGAPRANFKCN 120
DB 60 KEILSDINGIMKPGNALILGPTGGKSSLLDLVLAARKDPGLSDVLINGAPQANFKCS 119
QY 121 SGYVVQDDVVMGTLTVRENLOFSAAALRLATTTMTHNEKNERINRVIOELGLDKVADSKVGT 180
DB 120 SGYVVQDDVVMGTLTVRENLOFSAAALRLPKAMKTHEKNERINTIIEKELGLDKVADSKVGT 179
QY 181 QFIRCVSGGKRKRISIGNELITDPSILFLDEPTTGLDSSSTANAVLLLLKRMKSGKRTIIF 240
DB 180 QFTRGISGGERKRISIGNELITDPSILFLDEPTTGLDSSSTANAVLLLLKRMKSGKRTIIF 239
QY 241 SIHOPRYSIGFIKFDLSLTLLASGRMLMFGHPAQEALGYFESAGYHCEAYNNPADFLDIING 300
DB 240 SIHQPRYSIGFIKFDLSLTLLASGKMLMFGHPAQALEYFASAGYHCEPYNNPADFLDVIING 299
QY 301 DSTAVAIANR-EDFKATEIILPSKODKELIEKLAIEIVYNSFYKETAELHQLSGEKKK 359
DB 300 DSAVNLARGEODHAEKTEPSPKKEKIIENLAEFIYNSITTYGETKAELOQLPVAQKKK 359
QY 360 KITVFKEISYTTSFCHQLRWWSKSKFKNLGNPQASIAQIIVTVVLGVIGALIFGLKND 419
DB 360 GSSAFREPYYVTSFCHQLRWARRSFKNLGNPQASVLAQITVTVLGHIIGALYFGLKND 419
QY 420 STGICNIRAGVLFLLTNCFFSSVSAVELFVVEKKLFIHEYISGYRVSSYFGLKLLSLL 479
DB 420 PTGMQNRAGVFFLLTNCQFTSVSAVELFVVEKKLFIHEYISGYRVSSYFGLKLVSDLL 479
QY 480 PMTMLPSFIIFTCIVYFMILGLKEKADAFVMMFTLMMVYAGSSMAIAIAGQSVSVATL 539
DB 480 PMTMLPSFIIFTCIVYFMILGLKEKADAFVMMFTLMMVYAGSSMAIAIAGQSVSVATL 539

```

```

Db 480 PMRFLPSVIYTCILYFMLGLKRTVEAFFIMFTLIWVAYTASSMALAIAAGQSVSVATL 539
QY 540 LMTICFVFMIFSGLLVNLTTIASWLSWLOYSIPRYGFTALQHNELFGQFCGLNATG 599
Db 540 LMTISFVFMIFSGLLVNLRTIGPWSLWLOYSIPRYGFTALQHNELFGQFCGLNATG 599
QY 600 NNPC--NYATCTGEEYLVKQGDIDSPWGLWKNHVALACMIVIFLTIAYLKLFLKKYS 655
Db 600 NSTCVNSYITICTGNDYLIHQIDLSWGLWRNHNVALACMIIIFLTIAYLKLFLKKYS 657

RESULT 10
Q80XF3 ID Q80XF3 PRELIMINARY; PRT; 657 AA.
AC Q80XF3;
DT 01-JUN-2003 (T-REMBLrel. 24, Created)
DT 01-JUN-2003 (T-REMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (T-REMBLrel. 25, Last annotation update)
DE ATP-binding, cassette transporter ABCG2.
GN ABCG2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
SEQUENCE FROM N.A.
RA Shimano K., Satake M., Okaya A., Kitanaka J., Kitanaka N.,
RA Takemura M., Sakagami M., Terada N., Tsujimura T.;
RT "Hepatic Oval Cells Have the Side Population Phenotype Defined by
RT Expression of ATP-binding Cassette Transporter ABCG2/BCRP1.";
RL Am. J. Pathol. 0:0-0(2003).
DR EMBL; AB094089; BAC75866.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. .; IEA.
DR GO; GO:000166; P:nucleotide binding; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR006162; Ppantne.S.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
DR PROSITE; PS00012; PHOSPHOTRANSFERASE; 1.
KW ATP-binding.
SQ SEQUENCE 657 AA; 72961 MW; 459800CC3903DSCS CRC64;

Query Match 81.8%; Score 2742; DB 11; Length 657;
Best Local Similarity 80.9%; Pred. No. 1.1e-181;
Matches 532; Conservative 51; Mismatches 71; Indels 4; Gaps 3;

QY 1 MSSNVEVPIVPSQNTNGFPATASNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPVE 60
Db 1 MSSNDHVLVPMQSRNKLPGMSRGARTLAEDGVLSFHNITVRVKVSGFL-VRKTA 59
QY 61 KEILSNINGIMKPGNALIGTGGKSSLLDVLAAKDPGLSGDVLINGAPRANPKCN 120
Db 60 KEILSDINGIMKPGNALIGTGGKSSLLDVLAAKDPGLSGDVLINGAPRANPKCS 119
QY 121 SGYVQDDVVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIOELGDKVADSKVGT 180
Db 120 SGYVQDDVVMGTLTVRENLOFSAALRLFPKAMKTHEKNERINTIIEKLGDKVADSKVGT 179
QY 181 QPIRGVSGERKRTSIGMELITDPSILFDEPTTGLDSSSTANAVLLILKMSKQRTIIF 240
Db 180 QPTRIGSGERKRTSIGMELITDPSILFDEPTTGLDSSSTANAVLLILKMSKQRTIIF 239
QY 241 SIHQPRYSIFKLFDSLTLASGLMHPGPAQALGYFESAGYCHAEYNNPADPFLDLING 300
Db 240 SIHQPRYSIFKLFDSLTLASGLMHPGPAQALGYFESAGYCHAEYNNPADPFLDLING 299
QY 3012 DSTAVALNR-BEDFKATEIIEPSKQDKPLIEKLAIEIYVNSGPFYKTEKAEHLHQLSGGEKK 359

```

```

Db 300 DSSAVMLNRGEQHEANKTEEPSKREKPIENLAEYINSTIYGETKASDLQPLVQAKKK 359
QY 360 KITVKEISVTTTFCQHLRWVSKRSFKNLGNPQASIAQIIVTVLGLVIGATYFGLKND 419
Db 360 GSDFPEPVVTFCHQLRWIARRSKNLGNPQASVAQLIVTVLGLIIGALYFGLKND 419
QY 420 STGIQNRAGVFLFTTNQCPSSVAVELFVVEKLLFIHEYISGYRVSSYFGLKLSDLL 479
Db 420 PTGMQNRAGVFLFTTNQCPFTSVAVELFVVEKLLFIHEYISGYRVSSYFGLKLSDLL 479
QY 480 PMTMLPSIIFTCTIVYFMLGLKPKADAFFVNMFTLMVAVSASSMALAIAAGQSVSVATL 539
Db 480 PMRFLPSVIYTCILYFMLGLKRLVEAFFIMFTLIWVAYTASSMALAIAAGQSVSVATL 539
QY 540 LMTICFVFMIFSGLLVNLTTIASWLSWLOYSIPRYGFTALQHNELFGQFCGLNATG 599
Db 540 LMTISFVFMIFSGLLVNLRTIGPWSLWLOYSIPRYGFTALQHNELFGQFCGLNATG 599
QY 600 NNPC--NYATCTGEEYLVKQGDIDSPWGLWKNHVALACMIVIFLTIAYLKLFLKKYS 655
Db 600 NSTCVNSYITICTGNDYLIHQIDLSWGLWRNHNVALACMIIIFLTIAYLKLFLKKYS 657

RESULT 11
Q8BK15 ID Q8BK15 PRELIMINARY; PRT; 650 AA.
AC Q8BK15;
DT 01-MAR-2003 (T-REMBLrel. 23, Created)
DT 01-MAR-2003 (T-REMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (T-REMBLrel. 25, Last annotation update)
DE ATP-binding cassette.
GN ABCG3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RA STRAIN=C57BL/6J; TISSUE=Eye;
RC MEDLINE=22354683; PubMed=12466851;
RX The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RL Nature 420:563-573(2002).
DR EMBL; AK051880; BAC34799.1; -.
DR FIR; PT0645; PT0645.
DR MGD; MGI:1351624; Abcg3.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. .; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00005; ABC_tran; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
SQ SEQUENCE 650 AA; 73613 MW; B9395BDBDFD945E1 CRC64;

Query Match 53.3%; Score 1787.5; DB 11; Length 650;
Best Local Similarity 55.3%; Pred. No. 1.8e-115;
Matches 363; Conservative 104; Mismatches 174; Indels 15; Gaps 5;

QY 1 MSSNVEVPIVPSQNTNGFPATASNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPVE 60
Db 1 MASNNDPTVISMERHLCDLPTNTSDTLTTEZAVLSFHNISYQETVQSGFLPKKAV 60
QY 61 KEILSNINGIMKPGNALIGTGGKSSLLDVLAAKDPGLSGDVLINGAPRANPKCN 120
Db 61 IERLSNISIMKPGNALIMGPQDGRSLLDVLAAARDPGLSGDILINGKRPANFKCT 120
QY 121 SGYVQDDVVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIOELGDKVADSKVGT 180
Db 121 SGYVQDDVVMGTLTVRENLOFSAALRLPVTITRDEKRRINEVLELLHLNK-----E 173

```





"Full Length cDNA of gene At5g06530 (GI:15240083).";  
Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

[2]

SEQUENCE FROM N.A.  
Yanada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M., Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L., Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin X., Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Saito M., Seki M., Shinn P., Southwick A., Shinzaki K., Davis R.W., Ecker J.R., Theologis A.;  
"Arabidopsis Open Reading Frame (ORF) Clones."; Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
EMBL; AY059787; AAU24135.1; --  
EMBL; AY114068; AAM45116.1; --  
GO; GO:0016020; C:membrane; IEA.  
GO; GO:0005524; F:ATP binding; IEA.  
GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti...; IEA.  
GO; GO:0006810; P:transport; IEA.  
InterPro; IPR003439; ABC transporter.  
InterPro; IPR006162; Ppanne\_S.  
Pfam; PF00005; ABC\_tran; 1.  
DR ProDom; PD000006; ABC transporter; 1.  
DR PROSITE; PS00211; ABC TRANSPORTER\_1; 1.  
DR PROSITE; PS00893; ABC TRANSPORTER\_2; 1.  
DR PROSITE; PS00012; PHOSPHOTANETHEINE; 1.  
SQ SEQUENCE 751 AA; 82931 MW; 5564454F99D0A2EC CRC64;

Query Match 24.1%; Score 809; DB 10; Length 751;  
Best Local Similarity 33.9%; Pred. No. 1.7e-47;  
Matches 216; Conservative 110; Mismatches 248; Indels 66; Gaps 16;

QY 37 LSFNHCIVKVLKSGFLPCRPVEKEILSNINGIMPG-LNALIGPTGGKSKSLDLVIAA 95  
Db 157 LKFRDVTVKVIKK---LTSSVEKEILTGISGVNPEGLALMGPSGKGKTLLSILAG 212

QY 96 RDPGSLGDVLINGAPRPNPKNSGYVDVVMTLVRENLQFSAAURLATTWNH 155  
Db 213 RISQSSTGSVTYNKPSKYLUKIGFTQDDVLFPLTVEHLTYAARLPKTLTRE 272

QY 156 EXNERINRVIOELGDKVADSVKGTQPIRGVSGGERKRTSIGMELITDPSILFDDEPTTG 215  
Db 273 QXQRALDIVIOELGLERCQDTMIGAFVRGVSGGERKRVISGEIINIPSLLDDEPTSG 332

QY 216 LDSSITAVALLIKMSKOGRIIIFSIHQPRYSIFKLPSDLTLLASGRMLFHGAQAELG 275  
Db 333 LSDTTALRILMHDAEAGKVITTIHQPSRLFHRFDKILLORGSLLYFGKSSEALD 392

QY 276 YFESAGHYCEAYNNPADFFLDINDGSTAVALNREEDFXA-----TEIEBRSKDQP 327  
Db 393 YFSSIGCSPLIAMNPFAELFLDLANGINDISVPSELDDRQVGNSGRETQTGKPS--PAA 450

QY 328 LIEKLAELIYNSSFYKSTAEHLHOLS-GEEKKKITVFEK-----ISYTSFCHQLRW 379  
Db 451 VHEYLYEAETRVAEOEKKKLLDPVLPDDEAKAKSTRLKRWGTWCWQCYCILFCRGLKE 510

QY 380 VSKSRFKNLGNPQASIAQIIIVTVVLGLVIGAIYFGLKND----STGIQNRAGVLFPLTN 436  
Db 511 RRHEYFSWL-----RVTQVLSTAV---ILGLLW--QSDIRTPMGLQDAQLLFFIAVF 559

QY 437 QCFSV-SAVELFVVEKKLFIHEYISGYRVSYFGLKSLDLLPMPLPSIETCIYVF 495  
Db 560 WGFPPVPTAIFAPPQARMINKERAADMRLSAIFLARTSDLPDLDFILPS-LFLVLVVF 618

QY 496 MLGLKPKADAFFVMFTLMWVAYSASSMALATAAGOSVVSVALTMLTCFVPMIFSGLL 555  
Db 619 MTGLRISPYPFFLSMLTVFLCIIAAGLGAIGAILMDLKKATLASVTWTMFLAGGFF 678

QY 556 VNLTITASLWSLQYPSIPRYGTALQHNEFLQNFQCPGNATGNPNCAVTCBEYLAV 615  
Db 679 VKKPVP--FISWIRYLSFNTHTKLLKVKQY--QDFAVING-----MEI 719

QY 616 KQGIDLSFWGLWKNHVALACMIVIFLTAIYKLLFLX 652  
Db 720 DNGJ-----TEVALVVMIFYRLLYLVSLEQMK 748

RESULT 14  
Q9NH94 PRELIMINARY; PRT; 687 AA.  
ID Q9NH94  
AC Q9NH94  
DT 01-OCT-2000 (TRENBLrel. 15, Created)  
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
DE ATP dependent transmembrane transporter protein.  
GN WH3  
OS Bombyx mori (Silk moth).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Bombycoidea;  
OC Bombycidae; Bombyx.  
OX NCBI\_TaxID=7091;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Kin-Shiu X Sho-wa;  
RX MEDLINE=20469043; PubMed=11016828;  
RA Abraham E.G., Sezutsu H., Kanda T., Sugasaki T., Shimada T., Tamura T.;  
RT Identification and characterization of a silkworm ABC transporter gene homologous to Drosophila white.";  
RL Mol. Gen. Genet. 264:11-19(2000).  
CC 1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.  
DR EMBL; AF229609; AAF61569.1; --  
DR GO; GO:0016021; C:integral to membrane; IEA.  
DR GO; GO:0005524; F:ATP binding; IEA.  
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti...; IEA.  
DR GO; GO:000166; F:nucleotide binding; IEA.  
DR GO; GO:0006810; P:transport; IEA.  
DR InterPro; IPR003593; AAA ATPase.  
DR InterPro; IPR003439; ABC transporter.  
DR Pfam; PF00005; ABC tran; 1.  
DR ProDom; PD000006; ABC transporter; 1.  
DR SMART; SM00382; AAA; 1.  
DR TIGRFAM; TIGR00955; aa01204; 1.  
DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
DR PROSITE; PS00893; ABC\_TRANSPORTER\_2; 1.  
KW ATP-binding; Transmembrane; Transport.  
SQ SEQUENCE 687 AA; 75835 MW; ECD336333F0981AB CRC64;

Query Match 24.1%; Score 808; DB 5; Length 687;  
Best Local Similarity 32.5%; Pred. No. 1.7e-47;  
Matches 200; Conservative 108; Mismatches 248; Indels 60; Gaps 12

QY 61 KEILSNINGIMPG-LNALIGPTGGKSKSLDLVIAARKDPSGL--SGDVLINGAP-RPAN 116  
Db 100 KQLLRNVNGAAYPEGELLAIMGSGAGKTLLNTLTR-TFGGVATGTALNGQATPDA 158

QY 117 FKCN SGVWDVVGMGTLVRENLQFSAAURLATTWNHKNERNINRVIOELGDKVADS 176  
Db 159 LTALSAYVQQODLFITLVIREHLVFAQVMVMDRHIPYAQRMKRVQEIOLSKSCQT 218

QY 177 KVGTQ-FIRGVSGGERKRTSIGMELITDPSILFDDEPTTGLDSSITANAVLLLKMSKOG 235  
Db 219 VIGIPRLKGISGEMKRLSFASFVLDPLMFCDPTSGDSFMNQVIVLKLGAQWG 288

QY 236 RTIIFSHQPRYSIFKLPSDLTLLASGRMLFHGAQAELGYFESAGHYCEAYNNPADFFL 295  
Db 279 KTVVCTIHQPSSELYAMFKLLIMADGRVAFGLSSDEAQFFKELGAACPANTNPADHFI 338

QY 296 DINGOSTAVALNREDPFKATEIEBRSKQDKPLEKLAELIYNSSFYKSTAEHLHOLS 355  
Db 339 QLLAG-----VPGREEVTRHT-----IDTVCTAFKSEIGCRIAAEENALYN 381

QY 356 EKK-----KKITVFEKISYTTSPCHOLRWVSKSPFNKLLIGNPQASIAQIIVTV 403



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2004, 18:23:53 ; Search time 18 seconds  
(without alignments)  
1894.775 Million cell updates/sec

Title: US-09-961-086-1  
Perfect score: 3352  
Sequence: 1 MSSNVFVFPVQNGTNGF.....MIVIFLTAYLKLFLKKYS 655

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3339	99.6	655	1 ABG2 HUMAN	Q9UNQ0 homo sapien
2	1786.5	53.3	650	1 ABG3 MOUSE	Q99P81 mus musculus
3	835.5	24.9	1049	1 ADP1 YEAST	P25371 saccharomyc
4	812	24.2	687	1 WHIT DROME	P10090 drosophila
5	772	23.0	679	1 WHIT CERCA	Q17320 ceratitidis c
6	754.5	22.5	677	1 WHIT LUCCU	Q05360 lucilia cup
7	751	22.4	695	1 WHIT ANOGA	Q27256 anopheles g
8	736.5	22.0	709	1 WHIT ANOAL	Q16928 anopheles a
9	714.5	21.3	666	1 ABG1 MOUSE	C64343 mus musculus
10	708	21.1	678	1 ABG4 HUMAN	Q9H172 homo sapien
11	706.5	21.1	678	1 ABG1 HUMAN	P45844 homo sapien
12	694.5	20.7	652	1 ABG5 MOUSE	Q99P88 mus musculus
13	686.5	20.5	598	1 YPC3 CAHEL	Q11180 caenorhabdi
14	682.5	20.4	651	1 ABG5 HUMAN	Q9H222 homo sapien
15	682.5	20.4	651	1 ABG5 RAT	Q99P87 rattus norv
16	669.5	20.0	694	1 ABG8 RAT	P58428 rattus norv
17	665	19.8	673	1 ABG8 MOUSE	Q9DBM0 mus musculus
18	658.5	19.6	1294	1 YOH5 YEAST	Q08234 saccharomyc
19	654.5	19.5	666	1 SCRT DROME	P45843 drosophila
20	639.5	19.1	673	1 ABG8 HUMAN	Q9H221 homo sapien
21	627.5	18.7	610	1 YQ5C CAHEL	Q09466 caenorhabdi
22	612.5	18.3	1501	1 SNQ2 YEAST	P32568 saccharomyc
23	612	18.3	1530	1 BFRI SCHPO	P41820 schizosacch
24	602	18.0	1511	1 PD55 YEAST	P33302 saccharomyc
25	594	17.7	1564	1 PDRA YEAST	P51533 saccharomyc
26	593.5	17.7	1499	1 CDR2 CANAL	P78595 candida alb
27	585	17.5	1333	1 YN99 YEAST	P53756 saccharomyc
28	580.5	17.3	1333	1 YN99 YEAST	Q04182 saccharomyc
29	544	16.2	1511	1 PDRC YEAST	Q02785 saccharomyc
30	529	15.8	1501	1 CDR3 CANAL	O42690 candida alb
31	528	15.8	1501	1 CDR1 CANAL	P43071 candida alb
32	500	14.9	1490	1 CDR4 CANAL	O74676 candida alb
33	488.5	14.6	1410	1 PD8B YEAST	P40550 saccharomyc

34	474	14.1	675	1 BROW DROME	P12428 drosophila
35	455	13.6	668	1 BROW DROVI	Q24739 drosophila
36	243	7.2	1321	1 AB11 HUMAN	Q95342 homo sapien
37	242	7.2	355	1 CYSA_SNV3	P74548 synchocyst
38	240	7.2	338	1 CYSA_ANASP	Q82080 anabaena sp
39	238.5	7.1	1704	1 ABC3 HUMAN	Q99758 homo sapien
40	238	7.1	246	1 NATA_BACSU	P46903 bacillus su
41	235	7.0	371	1 MALK_ECOLI	P02914 escherichia
42	234.5	7.0	236	1 LIVF_ARCFU	C28882 archaeoglob
43	229.5	6.8	607	1 HEPA_ANASP	P22638 anabaena sp
44	229	6.8	1302	1 MDR4_DROME	Q00449 drosophila
45	228.5	6.8	362	1 CYSA_NITEU	Q82W5 nitrosomona

ALIGNMENTS

RESULT 1

ABG2 HUMAN STANDARD; PRT; 655 AA.

Q9UNQ0; Q95374; Q9BY73; Q9NUS0;

16-OCT-2001 (Rel. 40, Created)

28-FEB-2003 (Rel. 41, Last sequence update)

28-FEB-2003 (Rel. 41, Last annotation update)

DE ATP-binding cassette, sub-family G, member 2 (Placenta-specific ATP-binding cassette transporter) (Breast cancer resistance protein).

DE ABCG2 OR ABCP OR BCRP OR BCRP1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NCBI\_TaxID=9606;

[1]

SEQUENCE FROM N.A.

TISSUE=Placenta;

RX MEDLINE=99065313; PubMed=9850061;

RA Allikmets R., Schriml L.M., Hutchinson A., Romano-Spica V., Dean M.; "A human placenta-specific ATP-binding cassette gene (ABCP) on chromosome 4q22 that is involved in multidrug resistance.";

RT Cancer Res. 56:5337-5339(1998).

RL [2]

SEQUENCE FROM N.A.

RC TISSUE=Breast cancer;

RX MEDLINE=99080071; PubMed=9861027;

RA Doyle L.A., Yang W., Abruzzo L.V., Krogmann T., Gao Y., Rishi A.K., Ross D.D.; "A multidrug resistance transporter from human MCF-7 breast cancer cells.";

RT Proc. Natl. Acad. Sci. U.S.A. 95:15665-15670(1998).

RL [3]

ERRATUM.

RA Doyle L.A., Yang W., Abruzzo L.V., Krogmann T., Gao Y., Rishi A.K., Ross D.D.;

RL Proc. Natl. Acad. Sci. U.S.A. 96:2569-2569(1999).

RN [4]

SEQUENCE FROM N.A.

RP Kage K., Tsukahara S., Sugiyama T., Asada S., Ishikawa E., Tsuruo T., Sugimoto Y.;

RA "Breast cancer resistance protein constitutes a 140-kDa complex as a homodimer.";

RT Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.

RN [5]

SEQUENCE OF 198-655 FROM N.A.

RC TISSUE=Placenta;

RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y., Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H., Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M., Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S., Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K., Yamanoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Ninomiya K., Iwayanagi T.;

RT "NEDO human cDNA sequencing project.";

RT Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

RL [6]

RP REVIEW.  
 RX MEDLINE=21474438; PubMed=11590207;  
 RA Schmitz G., Langmann T., Heimerl S.;  
 RT "Role of ABCG1 and other ABCG family members in lipid metabolism.";  
 RL J. Lipid Res. 42:1513-1520(2001).  
 CC !- FUNCTION: Xenobiotic transporter that appears to play a major role  
 CC in the multidrug resistance phenotype of a specific MCF-7 breast  
 CC cancer cell line. When overexpressed, the transfected cells become  
 CC resistant to mitoxantrone, daunorubicin and doxorubicin, display  
 CC diminished intracellular accumulation of daunorubicin, and  
 CC manifest an ATP-dependent increase in the efflux of rhodamine 123.  
 CC !- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
 CC !- SIMILARITY: Belongs to the ABC transporter family. ABCG (White)  
 CC subfamily.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AF103796; AAC09188.1; -  
 CC EMBL; AF098951; AAC97367.1; -  
 CC EMBL; AB056867; BAB39212.1; -  
 CC EMBL; AK002040; BAA92050.1; -  
 CC Genew; HGNC:74; ABCG2.  
 CC MIM; 603756; -  
 CC GO; GO:0016021; C: integral to membrane; TAS.  
 CC GO; GO:0005524; P: ATP binding; TAS.  
 CC GO; GO:0004009; P: ATP-binding cassette (ABC) transporter acti. . .; TAS.  
 CC GO; GO:0005215; P: transporter activity; TAS.  
 CC GO; GO:0008559; P: xenobiotic-transporting ATPase activity; TAS.  
 CC GO; GO:0009315; P: drug resistance; TAS.  
 CC GO; GO:0006810; P: transport; TAS.  
 CC InterPro; IPR003593; AAA ATPase.  
 CC InterPro; IPR003439; ABC transporter.  
 CC Pfam; PF00005; ABC\_tran; 1.  
 CC ProDom; PD000006; ABC transporter; 1.  
 CC SMART; SM00382; AAA; 1.  
 CC PROSITE; PS00211; ABC\_TRANSPORTER\_1; FALSE\_NEG.  
 CC PROSITE; PS00893; ABC\_TRANSPORTER\_2; 1.  
 KW ATP-binding; Transmembrane; Transport.  
 FT DOMAIN 1 395 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 396 416 POTENTIAL.  
 FT DOMAIN 417 428 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 429 449 POTENTIAL.  
 FT DOMAIN 450 477 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 478 498 POTENTIAL.  
 FT DOMAIN 499 506 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 507 527 POTENTIAL.  
 FT DOMAIN 528 535 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 536 556 POTENTIAL.  
 FT DOMAIN 557 630 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 631 651 POTENTIAL.  
 FT DOMAIN 652 655 CYTOPLASMIC (POTENTIAL).  
 FT NP\_BIND 80 87 ATP (POTENTIAL).  
 FT CARBOHYD 418 418 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 557 557 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 596 596 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 24 24 V -> A (IN REF. 2 AND 4).  
 FT CONFLICT 166 166 E -> Q (IN REF. 1).  
 FT CONFLICT 208 208 F -> S (IN REF. 1).  
 FT CONFLICT 315 316 MISSING (IN REF. 5).  
 FT CONFLICT 482 482 R -> T (IN REF. 2).  
 SQ SEQUENCE 655 AA; 72343 MW; 89A6D351IDC5CCE0 CRC64;  
 Query Match 99.6%; Score 3339; DB 1; Length 655;  
 Best Local Similarity 99.5%; Pred. No. 6e-222;  
 Matches 652; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 MSSNVVEVFPVSQNTNGFPATVSNDLKAFTEGAVLSFHNICVRVKLSGFLPCRKPE 60

DB 1 MSSNVVEVFPVSQNTNGFPATVSNDLKAFTEGAVLSFHNICVRVKLSGFLPCRKPE 60  
 QY 61 KEILSNINGIMKPGINAILGPTGGKSSLLDLVLAARKDPSGLSDVLINGAPRANFKCN 120  
 DB 61 KEILSNINGIMKPGINAILGPTGGKSSLLDLVLAARKDPSGLSDVLINGAPRANFKCN 120  
 QY 121 SGYVQDDVVMGTITVRENLOFSAALRLATMTWHEKNERINRVIQELGLDKVADSVKGT 180  
 DB 121 SGYVQDDVVMGTITVRENLOFSAALRLATMTWHEKNERINRVIQELGLDKVADSVKGT 180  
 QY 181 QFIRGVSGERKRTSIGMELITDPSILFLDEPTTGLDSSSTANAVLLLLKRMKSGKRIIF 240  
 DB 181 QFIRGVSGERKRTSIGMELITDPSILFLDEPTTGLDSSSTANAVLLLLKRMKSGKRIIF 240  
 QY 241 SIHQPRYSIFKLFPSLTLLASGRIMRHEGPACALGYFESAGYHCEAYNNPADFFLDIING 300  
 DB 241 SIHQPRYSIFKLFPSLTLLASGRIMRHEGPACALGYFESAGYHCEAYNNPADFFLDIING 300  
 QY 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLABIYVNSPFYKTKAELHOLSGEKKKK 360  
 DB 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLABIYVNSPFYKTKAELHOLSGEKKKK 360  
 QY 361 ITVFKEISYTFSCHELRWVSKRSFKNLLGNPQASIAQIIIVTVLGLVIGAIYFGLKND 420  
 DB 361 ITVFKEISYTFSCHELRWVSKRSFKNLLGNPQASIAQIIIVTVLGLVIGAIYFGLKND 420  
 QY 421 TGIQNRAGVLPFLTTNQCFSSVSAVELEFVVEKGLFIHEYISGYRVSVSYFGLKLSDLLP 480  
 DB 421 TGIQNRAGVLPFLTTNQCFSSVSAVELEFVVEKGLFIHEYISGYRVSVSYFGLKLSDLLP 480  
 QY 481 MTWLSIFITFCIVYFMLGLKPKADAFVMMFTLMWVAYSASSMALAIAGOSVSVATLL 540  
 DB 481 MRMLPSIIFTCTIVYFMLGLKPKADAFVMMFTLMWVAYSASSMALAIAGOSVSVATLL 540  
 QY 541 MTICFVFMVIFSGLLVNLTTIASLSWLSQYFISIPRYGFTALQHNFFLGQNFQGLNATGN 600  
 DB 541 MTICFVFMVIFSGLLVNLTTIASLSWLSQYFISIPRYGFTALQHNFFLGQNFQGLNATGN 600  
 QY 601 NPCNVATCTGEEYLVKQIGIDLSFWGLKWKHVALACMIVIFLTIAVLLKLLFKKYS 655  
 DB 601 NPCNVATCTGEEYLVKQIGIDLSFWGLKWKHVALACMIVIFLTIAVLLKLLFKKYS 655  
 RESULT 2  
 ABG3\_MOUSE STANDARD; PRT; 650 AA..  
 ID ABG3\_MOUSE  
 AC Q99P81;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE ATP-binding cassette, sub-family G, member 3.  
 GN ABCG3.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 CX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Splesen;  
 RX MEDLINE=21030753; PubMed=11178751;  
 RA Mickleth L., Jain P., Miyake K., Schriml L.M., Rao K., Pojo T.,  
 RA Bates S., Dean M.;  
 RT "An ATP-binding cassette gene (ABCG3) closely related to the multidrug  
 RT transporter ABCG2 (MXR/ABCP) has an unusual ATP-binding domain.";  
 RL Mamm. Genome 12:86-88(2001).  
 CC !- SUBUNIT: May dimerize with another subunit to form a functional  
 CC transporter.  
 CC !- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
 CC !- TISSUE SPECIFICITY: Highest levels of expression in thymus and  
 CC spleen. Detected in lung and small intestine.  
 CC !- SIMILARITY: Belongs to the ABC transporter family. ABCG (White)  
 CC subfamily.

CC	-1- CAUTION: Seems to have a defective ATP-binding region.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>	
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	
CC	-----	
DR	EMBL; AF224242; AAK14241.1; -	
DR	MGI; 1351624; Abc3.	
DR	InterPro; IPR003439; ABC transporter.	
DR	Pfam; PF00005; ABC_tran; 1.	
DR	PROSITE; PS00211; ABC_TRANSPORTER_1; FALSE_NEG.	
DR	PROSITE; PSS0893; ABC_TRANSPORTER_2; 1.	
KW	Transmembrane; Transport.	
FT	DOMAIN 1 387	
FT	TRANSMEM 388 408	
FT	DOMAIN 409 420	
FT	TRANSMEM 421 441	
FT	DOMAIN 442 469	
FT	TRANSMEM 470 490	
FT	DOMAIN 491 498	
FT	TRANSMEM 499 519	
FT	DOMAIN 520 527	
FT	TRANSMEM 528 548	
FT	DOMAIN 549 623	
FT	TRANSMEM 624 644	
FT	DOMAIN 645 648	
FT	SEQUENCE 650 AA; 73623 MW; 86A5ABE4DD26971C CRC64;	
QY	Query Match	
DB	Best Local Similarity 53.3%; Score 1786.5; DB 1; Length 650;	
DB	Matches 363; Conservative 104; Mismatches 174; Indels 15; Gaps 5;	
QY	1 MSSNVEFIPVSGQNGTPATASNDLKAPTEGAVLSFENICVRVKLKGSLPCPKPVE 60	
DB	1 MASNNDPVTSMIRHELCDDPETNTSDLKUTEAEVLSFNFNISVQETVQSGFPLRKAYV 60	
QY	61 KEILSNINGIMKGNAILGTGGKSLDVLAAKDPGSLGSDVLINGAPRPNKCN 120	
DB	61 IERLSNIGIMKGNALMGQDGRSLLDVLAAARDPGLSGDILLINGKPRPNFKCT 120	
QY	121 SGYVQDDVNGTITVRENLOFSAAALRATMTNHEKNERINRVIQELGDKVADSKVGT 180	
DB	121 SGYVPQNDVLTGTVTRDNLEFSAALRPLVITRDEKRRRINEVLELHLNK-----E 173	
QY	181 QPIRVSGGERKRTSIGMELTDPISILFDEPTTGLDSSSTANAVLLLLKRMKSGQRTIIF 240	
DB	174 QNIKPSKELARKRSIAMELVTEHPILFDDPTTGLDLRTTDDVLVLRMSKKGRTIIF 233	
QY	241 SIHQPRYSIFKLFSLTLLASGRMLMHGPAQEAALGYFESAGYHCEAVNNPADPFLDLING 300	
DB	234 SINQPOYSIFKFFDSLTIVASGVKMFHGFQADALEYFRSAGYNESHNNPADPFLDVING 293	
QY	301 DSTAVALNRE---EDFKATEIIEPSKQDKPIELKLAELIYVNSVFYKTKAELHOLSGEK 357	
DB	294 GFSNILDTEEDGHEDDKTEELFERQY---VTGKLANNYAQSPLYSETRALDQLLGEQK 350	
QY	358 KKKITVFKIEISYTTSFCHQLRWVSRSPKNNLGNPQASIAQIIVTVVLGLVIGAIYGLK 417	
DB	351 LERSAV-ETTCVTFPCHQLKWIICQSPKPKGFVFWTVIQAIITVILATAVGTAFLYK 409	
QY	418 NDSTGIONRAGVLPFLTNQCFSSVAVELFVVEKKLFIHEYISGYRVSSVYFGKLLSD 477	
DB	410 NDCIEVQMRAGLLYLTTFQCITSVAGELFVIDRVRFHEHTSGYRVSSVYFGKLLAE 469	
QY	478 LLEPMTLPSIIFTCIVFMLGLKPKADAFVMMFTLMVAYSASMALAIAGOSVSVSA 537	
DB	470 LIPRLLPSTVSLITYVIAGVGMKCPFTMCTIMVLAYSASSLPLSIGAGENAVAP 529	
QY	538 TLLMTICFVFMWIFSGLLVNLTTIASWLSWLOQYFSIPRYGFTALQHNEFLGQFCPLNA 597	

Db	530 TLLVTIYFVFWLPSGLSLYPSGFLPKLSIQYFSPHYGFRALLHNEFLGQFCPEHNT 589
QY	598 TGNNPC-NYATCTGEEVLYKQGDLSLSPGLKKNHVALACMIVFLTIAYLKLLFLK 652
Db	590 EEVERCHNYVICTGEEFIMIQGIDLSWGFWENHLALVCTMILLITYVQLLOVK 645

RESULT 3

ID	ADP1_YEAST	STANDARD;	PRT;	1049 AA.
AC	P25371;			
DT	01-MAY-1992 (Rel. 22, Created)			
DT	01-MAY-1992 (Rel. 22, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Probable ATP-dependent permease precursor.			
GN	ADP1 OR YCR011C OR YCR11C OR YCR105.			
OS	Saccharomyces cerevisiae (Baker's Yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.			
OX	NCBI_TaxID=4932;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=92160395; PubMed=1789009;			
RA	Purnelle B, Skala J, Goffeau A.;			
RT	"The product of the YCR105 gene located on the chromosome III from			
RT	Saccharomyces cerevisiae presents homologies to ATP-dependent			
RT	permeases.";			
RL	Yeast 7:867-872 (1991).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=92327849; PubMed=1626432;			
RA	Skala J., Purnelle B., Goffeau A.;			
RT	"The complete sequence of a 10.8 kb segment distal of SUP2 on the			
RT	right arm of chromosome III from Saccharomyces cerevisiae reveals			
RT	seven open reading frames including the RVS161, ADP1 and PGK genes.";			
RL	Yeast 8:409-417 (1992).			
CC	-1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).			
CC	-1- SIMILARITY: Belongs to the ABC transporter family. MDR subfamily.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; X59720; CAA42328.1; -			
DR	PIR; S19421; S19421.			
DR	GermOnline; 138916; -			
DR	SGD; S0000604; ADP1.			
DR	GO; GO:0005783; C:Endoplasmic reticulum; IDA.			
DR	InterPro; IPR003593; AAA_ArPase.			
DR	InterPro; IPR003439; ABC_Transporter.			
DR	Pfam; PF00005; ABC_tran; 1.			
DR	ProDom; PD000006; ABC transporter; 1.			
DR	SMART; SM00382; AAA; 1.			
DR	PROSITE; PS00211; ABC_TRANSPORTER_1; 1.			
DR	PROSITE; PS00893; ABC_TRANSPORTER_2; 1.			
KW	ATP-binding; Transmembrane; Glycoprotein; Transport; Signal.			
FT	SIGNAL 1 25			
FT	CHAIN 26 1049			
FT	NP BIND 423 430			
FT	TRANSMEM 325 345			
FT	TRANSMEM 464 481			
FT	TRANSMEM 794 814			
FT	TRANSMEM 829 849			
FT	TRANSMEM 878 898			
FT	TRANSMEM 910 930			
FT	TRANSMEM 938 958			
FT	TRANSMEM 1001 1021			
FT	TRANSMEM 1025 1045			





CC cells responsible for eye color.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- SIMILARITY: Belongs to the ABC transporter family. MDR subfamily.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: X89933; CA61198.1; -  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC transporter.  
 DR InterPro; IPR005284; Pigment\_permease.  
 DR Pfam; PF00005; ABC\_tran; 1.  
 DR ProDom; PD000006; ABC\_transporter; 1.  
 DR SMART; SM00382; AAA; 1.  
 DR TIGRFAMs; TIGR00955; 3a01204; 1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
 DR PROSITE; PS00893; ABC\_TRANSPORTER\_2; 1.  
 KW Pigment; ATP-binding; Transmembrane; Transport.  
 FT NP BIND 121 128 ATP (BY SIMILARITY).  
 FT TRANSMEM 427 445 POTENTIAL.  
 FT TRANSMEM 457 477 POTENTIAL.  
 FT TRANSMEM 507 525 POTENTIAL.  
 FT TRANSMEM 534 555 POTENTIAL.  
 FT TRANSMEM 568 586 POTENTIAL.  
 FT TRANSMEM 651 670 POTENTIAL.  
 FT CARBOHYD 628 628 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 643 643 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 679 AA; 75145 MW; 3F9CB7C8A35C4C CRC64;

Query Match 23.0%; Score 772; DB 1; Length 679;  
 Best Local Similarity 30.7%; Pred. No. 1.8e-45;  
 Matches 209; Conservative 135; Mismatches 256; Indels 80; Gaps 21;

QY 11 PVSQNTN-GF-----PA-TASN-----DLKAFTEGAV----LSFHNICRYVK 47  
 DB 30 PYEQSSINGCFKNGYTLSPSPALTDNLTVSWNLDFV-GAVHQPSSGSKQLVNRVK 87  
 QY 48 ---LKSGLPCKPVEKILSNIMKPG-LNALGPTGGKSLDLVLAAR----- 97  
 DB 86 GVFCNERHIPAPR---KHLKNDGSAVYVPGELLAWSGGAGKTTLLNASAPRSKGVQI 144  
 QY 98 DPGSLSGDLVINGAPRA-NFKCNSGYVVDVVMGTLTVRENTQFSAALRLATMTNHE 156  
 DB 145 SFSTIR---MLNGHPVDAKEMQARCAVYQDDLFGLSLTAREHLIFQAVMRPHMTQKQ 201  
 QY 157 KNERINRVTOELGLKVDKADSKVGTQ-FIRGVSGGRKRTSTGMELITDPSILFDEPTTG 215  
 DB 202 KVRVDQVQLDLSLQKQNTLIGVPRVKGSLGGERKLAFASEALTDPPLLICDEPTSG 261  
 QY 216 LDSSTANAVLLKMSKGRITIIISIQPRYSIFKLPDLSLTLLASGLMFHGPQAQALG 275  
 DB 262 LDSFWAHSVQVVKLSQKGVILTIHQPSSELPFELDKILLMAEGRVALGTFGEAVD 321  
 QY 276 YFESAGHYCEAVNPADFLDIINGDSTAVALNREED-----FKATEIIEFSKQDKPLIE 330  
 DB 322 FFSYGATCPTNYTPADFVQVL-----AVVPEGEVESRDRVAKICDNFAVGKYSREMEQ 376  
 QY 331 KLAETVNSFSFKETKAEHLQSLGGKCKKITVFKIEISYTSFCHOLRWSKRSPFNLLG 390  
 DB 377 NQKLVKNSGFKQDE-----NEYTKASWNPQFRAVLWRSLVSK 418  
 QY 391 NPOASTAQIIVTVGLVITGALYFGLKNDSTGIONRAGVLFLTTNQCF-SVSAVELFV 449  
 DB 419 EPLLKVRLLQTTMAVLGLIFLQGLTQVGVWNINGAIFLFLTNMTFQNSPATITFT 478  
 QY 450 VEKKLPIFIYISGYRVSVSYFLGKLLSLLPLMTPLPSIIFTCTIVFMLGLPKADAFVM 509  
 DB 479 TELPVMRETRBLRCDFYFLQKTIAB-LPLFLVVPFLTAIYPLGLRPGVDHFFTA 537

QY 510 MFTLMVAYSASSMALAIAGQSVVSVATLMTICFVFMIFSGLLVNLTTIASWLSLQ 569  
 DB 538 LALVTLVANVSTFGYLISCACSSSTMSALSVGPVPIIPFLFGGFFLNSGSPVYFKWLS 597  
 QY 570 YSISRYGFTALQHNFEFLGQNFPCG-LNATGNPNVATCTGEEVLVKGGIDLSWGLWK 628  
 DB 598 YLSWRYANEGLLINQW--ADVKGEEITCTLSN-----TTCSSGEVILETINFSASDLFF 651  
 QY 629 NEVALACMIVIELTIAYLKL 648  
 DB 652 DFIGLALLIVGPRISAYIAL 671

## RESULT 6

WHIT\_LUCCU STANDARD; PRT; 677 AA.  
 ID AC Q05360;  
 DT 01-PEB-1995 (Rel. 31, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE White protein.  
 GN W.  
 OS Lucilia cuprina (Greenbottle fly) (Australian sheep blowfly).  
 CC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 CC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Oestroidea;  
 CC Calliphoridae; Lucilia.  
 CC NCBI\_TaxID=7375;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97087158; PubMed=8933176;  
 RA Garcia R.L., Perkins H.D., Howells A.J.;  
 RT "The structure, sequence and developmental pattern of expression of  
 RT the white gene in the blowfly *Lucilia cuprina*.";  
 RL Insect Mol. Biol. 5:251-260(1996).  
 RN [2]  
 RP SEQUENCE OF 490-584 FROM N.A.  
 RX MEDLINE=90264941; PubMed=1971656;  
 RA Elizur A., Vacek A.T., Howells A.J.;  
 RT "Cloning and characterization of the white and topaz eye color genes  
 RT from the sheep blowfly *Lucilia cuprina*.";  
 RL J. Mol. Evol. 30:347-358(1990).  
 CC -!- FUNCTION: May be part of a membrane-spanning permease system  
 CC necessary for the transport of pigment precursors into pigment  
 CC cells responsible for eye color.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- SIMILARITY: Belongs to the ABC transporter family. MDR subfamily.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: U38899; AAA82057.1; -  
 DB EMBL; X53265; CA37365.1; -  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC transporter.  
 DR InterPro; IPR005284; Pigment\_permease.  
 DR Pfam; PF00005; ABC\_tran; 1.  
 DR ProDom; PD000006; ABC\_transporter; 1.  
 DR SMART; SM00382; AAA; 1.  
 DR TIGRFAMs; TIGR00955; 3a01204; 1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
 DR PROSITE; PS00893; ABC\_TRANSPORTER\_2; 1.  
 KW Pigment; ATP-binding; Transmembrane; Transport.  
 FT NP BIND 119 126 ATP (POTENTIAL).  
 FT TRANSMEM 431 451 POTENTIAL.  
 FT TRANSMEM 456 476 POTENTIAL.  
 FT TRANSMEM 506 526 POTENTIAL.  
 FT TRANSMEM 534 554 POTENTIAL.



FT TRANSMEM 563 583 POTENTIAL.  
 FT TRANSMEM 647 667 POTENTIAL.  
 SQ SEQUENCE 677 AA; 75365 MW; D16FC11C97EED51D CRC64;  
 Query Match 22.5%; Score 754.5; DB 1; Length 677;  
 Best Local Similarity 31.4%; Pred. No. 2.9e-44;  
 Matches 194; Conservative 123; Mismatches 226; Indels 75; Gaps 18;

QY 5 NUVFTVPSQNTNGFPATASNDLKAFTEGAVLSFHNICVRVK---LKSGLPCRPVEK 61  
 DB 63 NLDVFEVHPGSEN-----WKQLNVRVKGVCNFRHHP--KP-RK 99  
 QY 62 EILSNINGIMKPG-LNAILGPTGGKSSLDVLAARK-----DPSGLSGDVLINGAPP 114  
 DB 100 HLIKCGVAYPGELLAVMGSSGAGKTLLNALAFRSARGVQISPSVR---MLNGHPVD 156  
 QY 115 A-NFKNSGVYVQDDVVMGTLTVRENLOFSALRLATTNHEKNERINRVQELGDKV 173  
 DB 157 AKEMQARCAVYQDDDLFIGSLTAREHLIQAATVVRMTQKQQRVDQVTDLSLKC 216  
 QY 174 ADSKVGTO-FIRGVSGERKRTSIGMELITDPSILFDEPTTGLDSTANAVALLKMS 232  
 DB 217 QNTIIGVPGRVKLSGGERKRLAFASEALTDPLLLCDEFTSLDSTFMAASVVQVLKLS 276  
 QY 233 KQRTIIFSIHQPRYSIFKLPDLSLTLASGRLMFHGPAQALGYFESAGHCAVNNPAD 292  
 DB 277 QRGKTIVLTTHQPSSELFELFDKILLMAEGRVAFGLTPVEAVDPFFIGAQCPNTNYPAD 336  
 QY 293 PFDIINGDSTAVALNREEDFKATEIIEPSKQDKPIEKLAETVYNSVFVKETKAELHOL 352  
 DB 337 FYVQVL-----AVPGRE-----IESDRISKIDCNFAVGKVSREMEQFQKI 379  
 QY 353 S-----GEEKKKKTIVFKEISYTSFCHQRWWSKRSFKNLLGNPQASIAQIIVTVGLV 408  
 DB 380 AAKTDGLQKDDETILYKASWFTQF----RAIMWRSWISTLKEPLLVKVLITQTMVAVL 435  
 QY 409 IGALYFGLKNDSTGIONRAGVLFELTNTQCFSSVAV-ELFVVEKKLFTHXISGYRYS 467  
 DB 436 IGLIFLNPQMTQGVWMINGAIFLFTNTWTFQNVFAVINFTSELVFRERSRLYRCD 495  
 QY 468 SYFLGKLLSLLPMTMLPSIIFTIVFVLMGLKPKADAFVMMFTLMMVAYSASSMALAI 527  
 DB 496 TYFLGKTLAE-LPLFLVPELFTAIAYPMIGLRPGITHLSALALVTLVANVSTSGYLI 554  
 QY 528 AAGOSVSVATLL---MTICFVFMIFSLGLLVNLTIASLWSLQVFSFRICFTALQHN 584  
 DB 555 SCATSTSMALSVPPLTIPTF---LLFGVFLNSGVVPVFKWLSYFSWFRYANEGLLIN 611  
 QY 585 EFL----GQNFPCPLNAT 598  
 DB 612 QWADVQPGEITCTSTNTT 629

RESULT 7  
 WHIT ANOGA STANDARD; PRT; 695 AA.  
 ID AC Q27256; Q17006;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE White protein.  
 GN W.  
 OS Anopheles gambiae (African malaria mosquito).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.  
 CX NCBI\_TaxID=7165;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Suakoko / G3;  
 RX MEDLINE=96423158; PubMed=8825759;  
 RA Besansky N.J., Bedell J.A., Benedict M.Q., Mukabayire O., Hilfiker D.,  
 RA Collins F.H.;  
 RT "Cloning and characterization of the white gene from Anopheles

gambiae.";  
 Insect Mol. Biol. 4:217-231(1995).  
 -!- FUNCTION: May be part of a membrane-spanning permease system necessary for the transport of pigment precursors into pigment cells responsible for eye color.  
 -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 -!- SIMILARITY: Belongs to the ABC transporter family. MDR subfamily.  
 -!- This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 EMBL; U29486; AAC46995.1; --  
 EMBL; U29485; AAC46994.1; --  
 EMBL; U29484; AAC47423.1; --  
 InterPro; IPR003593; AAA\_ATPase.  
 InterPro; IPR003439; ABC\_transporter.  
 InterPro; IPR008965; Cellul\_bind.  
 InterPro; IPR005284; Pigment\_permease.  
 Pfam; PF00005; ABC\_tran; 1.  
 ProDom; PD000006; ABC\_transporter; 1.  
 SMART; SM00382; AAA; 1.  
 TIGRFAMs; TIGR00955; 3a01204; 1.  
 PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
 PROSITE; PS00893; ABC\_TRANSPORTER\_2; 1.  
 KW Pigment; ATP-binding; Transmembrane; Transport.  
 NP\_BIND 133 140 ATP (POTENTIAL).  
 NP\_BIND 288 295 ATP (POTENTIAL).  
 TRANSMEM 444 464 POTENTIAL.  
 TRANSMEM 474 494 POTENTIAL.  
 TRANSMEM 524 544 POTENTIAL.  
 TRANSMEM 552 572 POTENTIAL.  
 TRANSMEM 581 601 POTENTIAL.  
 TRANSMEM 659 689 POTENTIAL.  
 CARBOHYD 472 472 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CONFLICT 100 100 N -> S (IN REF. 1; AAC47423).  
 CONFLICT 691 693 SRS -> YAR (IN REF. 1; AAC47423).  
 SQ SEQUENCE 695 AA; 77218 MW; ER8B9517239B2961 CRC64;  
 Query Match 22.4%; Score 751; DB 1; Length 695;  
 Best Local Similarity 30.0%; Pred. No. 5.3e-44;  
 Matches 206; Conservative 129; Mismatches 273; Indels 78; Gaps 20;

QY 6 VEVFIPVSQNTNG-FP-----ATASNDLKAFTEGAVLSFHNICVRVKLK 49  
 DB 50 IQVWPKSYGSVKGQIPQICERLTVYWEIDVFGAPTQKRP-REFLCTRLNCCTR--QR 106  
 QY 50 SGFLPCRPVEKEILSNINGIMKPG-LNAILGPTGGKSSLDVLAARKOPS-GLSGDVL 107  
 DB 107 KDFNP-----RRHLTKNVTGVAKSGELLAVMGSSGAGKTLLNALAFRSPGVKISPNV 161  
 QY 108 --INGAPRA-NFKNSGVYVQDDVVMGTLTVRENLOFSALRLATTNHEKNERINRV 164  
 DB 162 RALNGVPVNAEQLRARCAYVQDDLFIPSLTTRHELLFOAMLRMDRDPVPSVKCHRVQEV 221  
 QY 165 IQELGLDKVADSKVGTQ-FIRGVSGERKRTSIGMELITDPSILFDEPTTGLDSTANA 223  
 DB 222 IQELSLVKCADTIIGAPGRKGLSGGERKRLAFASETLTDPHLLCDEFTSLDSTFMAHS 281  
 QY 224 VILLILKMSKQRTIIFSIHQPRYSIFKLPDLSLTLASGRLMFHGPAQALGYFESAGYH 283  
 DB 282 VLQVLKGMWAKGKTIILTIHQPSSELYCLFDKILLVAEGRVAFGLSPYQSAEFFSQLGIP 341  
 QY 284 CEAVNNPADFELDINGDSTAVALNREEDFKATEIIEPSKQD--KPLIEKLAETVYNSVF 341  
 DB 342 CPPNYNPADFVQVLM-----AIAPAKEAECDRMKIKKICDSFAVSPI 382  
 QY 342 YKETKAELHQLSGGEKK----KKITVPKEISYTSFCHQLRWWSKRSFKNLLGNPQASIA 397

383 AREVLETSVAGKMGDEPYMLQOEVGSGTGRSSWMTQFCILWRGWSLVKOPMLVKV 442  
 398 QIIVTVVLGLVIGALYFGLKNDSTGICQNRAGVLFELTNQCFSSVSVA-ELFVVEKKLFI 456  
 443 RLQQTAMVATLIGSYFQGVLDQGVNINGSFLFELTNMTFFQNVFAVINVSABLPVFL 502  
 457 HEIYISGYRVSSYFLGKLLSDLLPMTLPSIIFTCIVYFVLMGLPKADAFVMMFTLMV 516  
 503 REKRSRLYRVDYFLGKTIAB-LPLFIAPVFTSITVPMIGLRTGATHYTLTLFIVTLV 561  
 517 AYSASSMALATAAGOSVSVATLLMTICFVPMIFSGLLVNLTTIASNLSWLOFSPRY 576  
 562 ANVSTFGLYSCASSISMSALSGVPPVIFLIFGGFFLNSASVPAFYKLSYLSWFRY 621  
 577 GFTALQHNEL----GQNFPCGLNATGNNPCNYATCTGEEYLKQ-GIDLSPMGLKXHV 631  
 622 ANEALLINQWSTVVDGEIACTRAN-----VTCPESEIILETFNFRVEDPAL----- 667  
 632 ALACM--IVIFLTIAYLKLFKKYS 655  
 668 DIACLFALIVLFRGALCLMLRSRS 693

RESULT 8  
 ID WHIT\_ANOAL STANDARD; PRT; 709 AA.  
 AC Q16928;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE White protein.  
 GN W.  
 OS Anopheles albimanus (New world malaria mosquito).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.  
 OX NCBI\_TaxID=7167;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN:TeCo.  
 RX MEDLINE=9826549; PubMed=9465401;  
 RA Ke Z., Benedict M.Q., Cornet A.J., Besansky N.J., Collins F.H.;  
 RT "The Anopheles albimanus white gene: molecular characterization of  
 the gene and a spontaneous white gene mutation."  
 RL Genetics 101:87-96(1997).  
 CC -!- FUNCTION: May be part of a membrane-spanning permease system  
 CC necessary for the transport of pigment precursors into pigment  
 CC cells responsible for eye color.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- SIMILARITY: Belongs to the ABC transporter family. MDR subfamily.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; L76302; AAA88240.1; -.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR InterPro; IPR005284; Pigment\_permease.  
 DR Pfam; PF00005; ABC\_tran; 1.  
 DR ProDom; PD000006; ABC\_transporter; 1.  
 DR SMART; SM00382; AAA; 1.  
 DR TIGRFAMs; TIGR00955; 3a01204; 1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; FALSE\_NEG.  
 DR PROSITE; PS00893; ABC\_TRANSPORTER\_2; 1.  
 KW Pigment; ATP-binding. Transmembrane; Transport.  
 FT NP\_BIND 136 143 ATP (POTENTIAL).  
 FT NP\_BIND 292 299 ATP (POTENTIAL).  
 FT TRANSMEM 457 475 POTENTIAL.

FT TRANSMEM 487 507 POTENTIAL.  
 FT TRANSMEM 537 555 POTENTIAL.  
 FT TRANSMEM 564 585 POTENTIAL.  
 FT TRANSMEM 598 616 POTENTIAL.  
 FT TRANSMEM 681 700 POTENTIAL.  
 FT CARBOHYD 485 485 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 658 658 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 709 AA; 79052 MW; FBD76D5C69D3BECA CRC64;  
 Query Match 22.0%; Score 736.5; DB 1; Length 709;  
 Best Local Similarity 30.1%; Pred. No. 5.4e-43;  
 Matches 210; Conservative 135; Mismatches 265; Indels 87; Gaps 24;  
 QY 6 VEVFIPVSGNTNG-FPAT-----ASNDLKAPTEGAV--LSFHNICRYV-----KLKS 50  
 DB 51 IQWRPKSYGSYKVGQIPAQDELTYTWREIDVFGAAIDKGRPLCSLRHCFTRQRLVK 110  
 QY 51 GLPLCKPCKVEKILNININGIMKPG-LNAILGPTGGKSKSLDVLAAARKDPS-GLSGDVL- 107  
 DB 111 DFNP-----RKHLKXVTVGARSSELLAVMGSSGAGKTTLLNELAFRPPGVKISPNAR 165  
 QY 108 -INGAPRPA-NPKCMSGVVQDDVVMGTLTVRENIQPSAALRL-----ATTNHEKNER 160  
 DB 166 TLNGVPTVTAEQMRARCAVYQDDLFIPSLTTKEHLMFQAMLEMGDRDVFATPKMH---R 221  
 QY 161 INRVQELGLDVADSKVGTQ-FIRGVSGGERKRTSIGMELITDPSIILFDEPTTGLDSS 219  
 DB 222 VDEVLQELSLVKADTIIGVAGRVKLSGGERKRTAFRSEITLDPHLLCDEPTSLDSF 281  
 QY 220 TANAVLLLLKRMKQGRITFIISIHOPRYSIFKLPDLTLASGRMLMFGHGAALGYFES 279  
 DB 282 MAQSVLQVLKGMAMKGTIILTIHQPSSELYCLFDRILLVAEG-VAFLGSPYQSADFFSQ 340  
 QY 280 AGVHCAYNPNADPFLDIINGSTVALNREDEKATEIIEPSKQDKPLIEKLAELVYNS 339  
 DB 341 LGIPCPNPVNPADFYVQML-----ALAPNKETERET-----IKKICDSFAVS 383  
 QY 340 SFYK---ETKAEHLQLSGGE-KKKKITV-----FKEISYTTSFCHQLRWYSKRSF 385  
 DB 384 PIARDIETASQVNGDGGIELTRKHTTDPYFLQPMEGVDSTGYRASWMTQFCILWRSM 443  
 QY 386 XNLLGNPQASIAQIVTVVLGVLGNAIVEGLKNDSTGIONRAGVLFLTNQCFSSVSVA 445  
 DB 444 LSVLKDPMLVKVRLQLQTAVASLIGSYFQGVLDQGVNINGSFLFELTNMTFFQNVFA 503  
 QY 446 -ELFVVEKKLFIHEIYISGYRVSSYFLGKLLSDLLPMTLPSIIFTCIVYFVLMGLPKAD 504  
 DB 504 INVFSALPFLREKESRLYRVDYFLGKTIAB-LPLFIAPVFTSITVPMIGLKAITS 562  
 QY 505 APTVMFTLMVAYSASSMALAIAAGOSVSVATLLMTICFVPMIFSGLLVNLTTIASN 564  
 DB 563 HYLTTILFIVTLVANVSTSFGLYSCASSISMSALSGVPPVIFLIFGGFFLNSASVPAY 622  
 QY 565 LSMLOVFSIPRYGFTALQHNEL----GQNFPCGLNATGNNPCNYATCTGEEYLKQGD 620  
 DB 623 FKLSYLSWFRYANEALLINQWADHEDGEIGCTRANVT-----CPASGEIILETFNER 675  
 QY 621 LSPWGLWQKHVALACM--IVIFLTIAYLKLFKKYS 655  
 DB 676 VEDFAL-----DIGCLFALIVLFRGALCLMLRSRS 707

RESULT 9  
 ABGI\_MOUSE STANDARD; PRT; 666 AA.  
 ID ABGI\_MOUSE  
 AC Q64343;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE ATP-binding cassette, sub-family G, member 1 (White protein homolog)  
 DE (ATP-binding cassette transporter 8).  
 GN ABCG1 OR ABCG8 OR WH1L.  
 OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97186700; PubMed=9034316;  
 RA Croop J.M., Tiller G.E., Fletcher J.A., Lux M.L., Raab E.,  
 RA Goldenson D., Son D., Arciniegas S., Wu R.,  
 RT "Isolation and characterization of a mammalian homolog of the  
 RT Drosophila white gene";  
 RL Gene 185:77-85(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DBA/2;  
 RX MEDLINE=96359154; PubMed=87031120;  
 RA Savary S., Denizot F., Luciani M.-F., Mattei M.-G., Chimini G.,  
 RT "Molecular cloning of a mammalian ABC transporter homologous to  
 RT Drosophila white gene";  
 RL Mamm. Genome 7:673-676(1996).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21092576; PubMed=11162488;  
 RA Lorkowski S., Rust S., Engel T., Jung E., Tegeltkamp K., Galinski E.A.,  
 RA Asmann G., Cullen P.,  
 RT "Genomic sequence and structure of the human ABCG1 (ABC8) gene";  
 RL Biochem. Biophys. Res. Commun. 280:121-131(2001).  
 RN [4]  
 RP INDUCTION, AND PROBABLE FUNCTION.  
 RX MEDLINE=20261604; PubMed=10799558;  
 RA Venkateswaran A., Repa J.J., Lobaccaro J.-M.A., Bronson A.,  
 RA Mangelsdorf D.J., Edwards P.A.,  
 RT "Human white/murine ABC8 mRNA levels are highly induced in  
 RT lipid-loaded macrophages. A transcriptional role for specific  
 RT oxysterols";  
 RL J. Biol. Chem. 275:14700-14707(2000).  
 RN [5]  
 RP REVIEW.  
 RX MEDLINE=21474438; PubMed=11590207;  
 RA Schmitz G., Langmann T., Heimerl S.,  
 RT "Role of ABCG1 and other ABCG family members in lipid metabolism.";  
 RL J. Lipid Res. 42:1513-1520(2001).  
 CC -I- FUNCTION: Transporter involved in macrophage lipid homeostasis. Is  
 CC an active component of the macrophage lipid export complex. Could  
 CC also be involved in intracellular lipid transport processes. The  
 CC role in cellular lipid homeostasis may not be limited to  
 CC macrophages.  
 CC -I- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
 CC -I- TISSUE SPECIFICITY: Expressed mainly in brain, thymus, lung,  
 CC adrenal, spleen and placenta. Little or no expression in liver,  
 CC kidney, heart, muscle or testes.  
 CC -I- INDUCTION: Strongly induced in macrophage cell line RAW264.7  
 CC during cholesterol influx. Induction is mediated by the liver X  
 CC receptor/retinoid X receptor (LXR/RXR) pathway.  
 CC -I- SIMILARITY: Belongs to the ABC transporter family. ABCG (white)  
 CC subfamily.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U34920; AAB47738.1; --  
 CC EMBL; Z49745; CAA88636.1; --  
 CC EMBL; AF323659; AAK27442.1; --  
 CC MGD; MGI:107704; Abcg1.  
 CC InterPro; IPR003593; AAA\_ATPase.  
 CC InterPro; IPR003439; ABC\_transporter.  
 CC InterPro; IPR005284; Pigment\_permease.  
 CC Pfam; PF00005; ABC\_tran; 1.  
 CC ProDom; PD000006; ABC\_transporter; 1.

DR SMART: SMO0382; AAA: 1.  
 DR TIGSPAMS; TIGR00955; J01204; 1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
 DR PROSITE; PS00893; ABC\_TRANSPORTER\_2; 1.  
 KW Transport; Lipid transport; ATP-binding; Transmembrane.  
 FT DOMAIN 1 414 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 415 433 EXTRACELLULAR (POTENTIAL).  
 FT DOMAIN 434 444 POTENTIAL.  
 FT TRANSMEM 445 465 POTENTIAL.  
 FT DOMAIN 466 494 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 495 513 POTENTIAL.  
 FT DOMAIN 514 521 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 522 543 POTENTIAL.  
 FT DOMAIN 544 555 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 556 574 POTENTIAL.  
 FT DOMAIN 575 637 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 638 657 POTENTIAL.  
 FT DOMAIN 658 666 CYTOPLASMIC (POTENTIAL).  
 FT NP\_BIND 118 125 ATP (POTENTIAL).  
 SQ SEQUENCE 666 AA; 74033 MW; EDDC6AFBD43950B6 CRC64;  
 Query Match 21.3%; Score 714.5; DB 1; Length 666;  
 Best Local Similarity 28.9%; Pred. No. 1.6e-41;  
 Matches 198; Conservative 145; Mismatches 244; Indels 97; Gaps 24;  
 QY 3 SSVVEVPIVSOQNTNGFPATASNDLKAFTE-----GAV-LSFHNICVRVK---- 47  
 DB 38 SSVVD---EVEITLLNGHLKKVDNN---FTEAQRFSLPRAAVNTEFKDLSYVPEGPW 91  
 QY 48 -LKSGLPCRKPVEKILSNINGIMKPG-LNALIGPTGGGKSLDVLAAKDPGSLSGD 105  
 DB 92 WKXGKY-----KTLKLGISGKNSGELVALNGPSGACKSTLMNLAYRE-TGMKGA 142  
 QY 106 VLINGAPRAN-PKCSGYVQDDVVGVTITVENLOFSALRLATTTWTHEKNERINRV 164  
 DB 143 VLINGMPRLRCERKVCYIMQDDMLPLHTVQEAAMVSAHLKQE--KDEGRREVKYEI 200  
 QY 165 IQELGLDKVADKVGTOFIRGVSGGERKRTSIGNELITDPSILFLDEPTTGLDSTANAV 224  
 DB 201 LTALGLPLCANTRTGS-----LSGQQRKRLAIALELVNNPVPWFDEFTSGLDSASCFQV 255  
 QY 225 LLLAKRMSKQRTIIFSIHQPRYSIFKLFDSLTLASGRMLFHGPAQEAALGYFESAGVHC 284  
 DB 256 VSLMKGLAQGQSGIVCTIHPQSAKLFELFDQLYVLSQGCYVRGKVSNLVPLRLDLGLNC 315  
 QY 285 EAYNPADEFLLIINGP-----STAVALNRE-----EDFKATEIIEPSKQDKPIEK 331  
 DB 316 PTYNPADFVNEVASGEYGDQNSRLVRAVREGMCDADYKDLGGDTDVNFFLWHPAEED 375  
 QY 332 LAEIVVNSSFYKETKAEHLQLSGGKKKKTIVFKEISYTTTSFCHQLRWKSKRSFKNLGN 391  
 DB 376 SASMEGCHSF-----SASCLTQFC-----ILPKRTFLSIMRD 407  
 QY 392 PQASIAQITIVTVGLVIGAIYFGLKNDSTGIQNRAGVLPFLTNQCFSSVSVEL-FVV 450  
 DB 408 SVLTHLRITSHIGLGLIGLDYLGIGNEAKKVLNSGFLFSLFMLFPAALMPTVLPPL 467  
 QY 451 EKCLFIHEYISGYRVSSYFLGKLLSDLLPMTLPSIIFTCIVYFMLGLPKADAFVPM 510  
 DB 468 EMSVFLREHLNVYSLKAYLAKTMAD-VPEQIMFPVAYGSIYVWMTS-QPSDAVRFLVF 525  
 QY 511 FTL-MWVAYSASSMALAIACQSVSVATLMTCTCFVMMIFPSGLLNLITIASLWLSLQ 569  
 DB 526 AALGTMTSLSVAQSLGLIGLAASTSLQVATFVGPTAIFVLLFSGFFVFDFTIPAYLQMS 585  
 QY 570 YFSIPRYGFTALQHNELFQNGFCPLNATGNPNPNYA-TC--TGEELVKGQIDLSPWGL 626  
 DB 586 YISVRYGFEVILSIY-----GLDREDLH-CDIAETCHFKQSEAILRE-LDVENAKL 636  
 QY 627 WKNHVALACMTVIFETAYLKLFF 650  
 DB 637 YLDFIVLGIFFISLRIAYFVIRY 660

## RESULT 10

ABCG4 HUMAN  
ID ABCG4\_HUMAN STANDARD; PRT; 646 AA.  
AC Q9HJ72;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE ATP-binding cassette, sub-family G, member 4.  
GN ABCG4 OR WHITE2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21518231; PubMed=11606068;  
RA Engel T., Lorkowski S., Lueken A., Rust S., Schluster B., Berger G.,  
Cullen P., Assmann G.;  
RT "The human ABCG4 gene is regulated by oxysterols and retinoids in  
monocyte-derived macrophages.";  
RL Biochem. Biophys. Res. Commun. 288:483-488(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg K.L., Feingold E.A., Grouse L.H., Derge J.G.,  
Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
Brownstein M.J., Udwin T.B., Toshynki S., Carninci P., Prange C.,  
Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Rulyk S.W.,  
Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
Fahey J., Helton E., Kettner M., Madan A.C., Rodriguez S., Sanchez A.,  
Whiting M., Madan A.C., Young A.C., Shevchenko Y., Bouffard G.G.,  
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
Butterfield Y.S.N., Krzywinski M.J., Skalska U., Smallus D.E.,  
Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [3]  
RP SEQUENCE OF 20-646 FROM N.A.  
RC TISSUE=Dorsal root ganglion;  
RX MEDLINE=22170423; PubMed=12183068;  
RA Oldfield S., Lowry C., Ruddick J., Lightman S.;  
RT "ABCG4: a novel human white family ABC-transporter expressed in the  
brain and eye.";  
RL Biochim. Biophys. Acta 1591:175-179(2002).  
CC !- FUNCTION: May be involved in macrophage lipid homeostasis.  
CC !- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
CC !- SIMILARITY: Belongs to the ABC transporter family. ABCG (White)  
subfamily.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/annouce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC  
CC EMBL; AJ308237; CAC87131.1; -;  
CC EMBL; BC041091; AAH41091.1; -;  
CC EMBL; AJ300465; CAC17140.1; -;  
CC FIR; JC7777; JC7777.  
CC Genew; HGNC:13884; ABCG4.  
CC MIM; 607784; -;

DR InterPro; IPR003593; AAA\_ATPase.  
DR InterPro; IPR003439; ABC\_transporter.  
DR Pfam; PF00005; ABC\_tran; 1  
DR ProDom; PD000006; ABC\_transporter; 1.  
DR SMART; SM00382; AAA; 1.  
DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
DR PROSITE; PS0093; ABC\_TRANSPORTER\_2; 1.  
KW ATP-binding; Glycoprotein; Transmembrane; Transport.  
FT DOMAIN 1 393  
FT TRANSMEM 394 414  
FT TRANSMEM 415 425  
FT TRANSMEM 426 446  
FT TRANSMEM 447 472  
FT TRANSMEM 473 493  
FT TRANSMEM 494 503  
FT TRANSMEM 504 524  
FT TRANSMEM 525 532  
FT TRANSMEM 533 553  
FT TRANSMEM 554 617  
FT TRANSMEM 618 636  
FT TRANSMEM 639 646  
FT NP\_BIND 102 109  
FT CARBOHYD 422 422  
SQ SEQUENCE 646 AA; 71895 MW; 9CC6C6E150772611 CRC64;  
Query Match 21.1%; Score 708; DB 1; Length 646;  
Best Local Similarity 27.7%; Pred. No. 4.4e-41;  
Matches 176; Conservative 141; Mismatches 247; Indels 72; Gaps 16;  
Qy 37 LSPHNTCYRVKLSGFLPC-RKPEVEKEILSNING-IMKPLGNAILGPTGGKSSLLDLVA 94  
Db 61 IEFVELSV--REG-PCWRKRGVYKTLKLSGKFCRELIGINGPSGAGKSTFNILA 116  
Qy 95 ARKDPGLSGDVLINGAPRP-ANFKCNSGVYVDDVWGTLVRENLOESALRLATYMT 153  
Db 117 GYRE-SGMGQILVNGRPRLTFRMOSCYINQDDMLPLHLTVLEAMVMSAKLUSE--K 173  
Qy 154 NHEKNERINRVIQELGLDGVADSKVGTQIRGVSGGERKRTSIGMELITDPSILFDEPT 213  
Db 174 QEVKELVTEILTALGLMSCSHTRTAL-----LSGGQRKLAIALELVNPPVWFDEPT 228  
Qy 214 TGLDSTANAVALLKRMKQGRITIFSIHQPRYIFKLFDSLTLLASRLMFHQAQA 273  
Db 229 SGDSASCQVYVSLMKSLAQGGRTICTTHQPSAKLFEMFDKLYILSQGQCIFKGVVNL 288  
Qy 274 LGYFESAGYHCAYNNPADFFLDIINGD-----STAVALNREEDFKATEII 319  
Db 289 IFLKGLGLHCTYHNPADFFIEVASGEYGLNPLMFRVQNGLCMAEKSSPEKNEVP 348  
Qy 320 EPSKQDKPLIEKLAIIYVNSPYKETKAEHLHOLSGEKKKKTIVPKKISYTTSPCHQLRW 379  
Db 349 APCPPCPPEVDPI-----ESH-----TFATSTLTQFCI 376  
Qy 380 VSKRSFKNLIGNPQASIAQIIYTVVLGLVIGAIYFGLKNDSTGIONRAGVLFELTNQCF 439  
Db 377 LFKRTFELSDRLDTVLTRFMSHVIGVLIGLYLHIGDDASKVFNNTGCLFSLMLFMF 436  
Qy 440 SSVSAVEL-FVVEKKLFTIHEYISGYRVSSYFLGKLSLDLPMTPMLPSIIFTCIVYFMG 498  
Db 437 AALMPTVLTPLPLEMAVEMREHLNLYSLKAYVLAKTMAD-VPFQVVCVYVVCISVYVMTG 495  
Qy 499 LKPKADAFVMMFTLMMVAYSASSMALATAAGQSVSVATLMTICFVFMMLFSGLLVNL 558  
Db 496 QPAETSRFLFSAATATATVALAQSGLLIGASNSLOVATFVGPVTAIPVLVLSGFFVSF 555  
Qy 559 TTITASLWMLQVPSIPRYGFTALQHNFEFLGQFCPLNATGNPNPNYATCT-GEYLVKQ 617  
Db 556 KTIPTYLQSSYLSYVYVGFEGVILTY-----GME-RGDLTCLBERCPPEQSILR 607  
Qy 618 GIDLSFAGLWQKHVALACMIVIFLTAYLKLFLKK 653  
Db 608 ALDVEDAKLYMDFLVLGIFFLALRLALYLVLRYRK 643

```

RESULT 11
ID ABG1_HUMAN STANDARD; PRT; 678 AA.
AC P45844; Q9BXK6; Q9BXK7; Q9BXK8; Q9BXK9; Q9BXL0; Q9BXL1; Q9BXL2;
AC Q9BXL3; Q9BXL4;
DT 01-NOV-1995 (Rel. 32, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE ATP-binding cassette, sub-family G, member 1 (White protein homolog)
DE (ATP-binding cassette transporter 8).
DE ABCG1 OR ABCS OR WH1.
GN ABCG1 OR ABCS OR WH1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE=20289799; PubMed=10830953;
RC TISSUE=Retina;
RX MEDLINE=96256850; PubMed=8659545;
RA Chen H.M., Rossier C., Laliet M.D., Lynn A., Chakravarti A.,
RA Perrin G., Antonarakis S.E.;
RA "Cloning of the cDNA for a human homologue of the Drosophila white
RT gene and mapping to chromosome 21q22.3.";
RL Am. J. Hum. Genet. 59:66-75(1996).
RN [2]
RP MEDLINE=20289799; PubMed=10830953;
RX MEDLINE=20289799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Groner Y.,
RA Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K.,
RA Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA Minoshima S., Shimizu N., Nordiek G., Hornischer K., Brandt P.,
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
RA Ramser J., Beck A., Klages S., Hennig S., Rieselmann L., Dgand E.,
RA Wehmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA Lehrach H., Reinhardt R., Yaspo M.-L.;
RA "The DNA sequence of human chromosome 21.";
RT Nature 405:311-319(2000).
RN [3]
RP MEDLINE=20408883; PubMed=10950923;
RX MEDLINE=20408883; PubMed=10950923;
RA Berry A., Scott H.S., Kudoh J., Talior I., Korostishevsky M.,
RA Wattenhofer M., Guipponi M., Barras C., Rossier C., Shibuya K.,
RA Wang J., Kawasaki K., Asakawa S., Minoshima S., Shimizu N.,
RA Antonarakis S.E., Bonne-Tamir B.;
RT "Refined localization of autosomal recessive nonsyndromic deafness
RT DFNB10 locus using 34 novel microsatellite markers, genomic
RT structure, and exclusion of six known genes in the region.";
RL Genomics 68:22-29(2000).
RN [4]
RP MEDLINE=21192304; PubMed=11279031;
RX MEDLINE=21192304; PubMed=11279031;
RA Porsch-Oezcuermes M., Langmann T., Heimerl S., Borsukova H.,
RA Kaminski W.E., Dobnik W., Honer C., Schumacher C., Schmitz G.;
RT "The zinc finger protein 202 (ZNF202) is a transcriptional repressor
RT of ATP binding cassette transporter A1 (ABCA1) and ABCG1 gene
RT expression and a modulator of cellular lipid efflux.";
RL J. Biol. Chem. 276:12427-12433(2001).
RN [5]
RP MEDLINE=21092576; PubMed=11624488;
RX MEDLINE=21092576; PubMed=11624488;
RA Lorkowski S., Rust S., Engel T., Jung E., Tegelkamp K., Galinski E.A.,
RA Asmann G., Cullen P.;
RT "Genomic sequence and structure of the human ABCG1 (ABC8) gene.";
RL Biochem. Biophys. Res. Commun. 280:121-131(2001).
RN [6]
RP MEDLINE=97186700; PubMed=9034316;
RX MEDLINE=97186700; PubMed=9034316;
RA Korkowski S., Rust S., Engel T., Jung E., Tegelkamp K., Galinski E.A.,
RA Asmann G., Cullen P.;
RT "Genomic sequence and structure of the human ABCG1 (ABC8) gene.";
RL Biochem. Biophys. Res. Commun. 280:121-131(2001).
RN [7]
RP MEDLINE=20261604; PubMed=10799558;
RX MEDLINE=20261604; PubMed=10799558;
RA Venkateswarar A., Repa J.J., Lobaccaro J.-M.A., Bronson A.,
RA Mangelsdorf D.J., Edwards P.A.;
RT "Human white/murine ABC8 mRNA levels are highly induced in
RT lipid-loaded macrophages. A transcriptional role for specific
RT oxysterols.";
RL J. Biol. Chem. 275:14700-14707(2000).
RN [8]
RP MEDLINE=20105556; PubMed=10639163;
RX MEDLINE=20105556; PubMed=10639163;
RA Klucken J., Buechler C., Orso E., Kaminski W.E.,
RA Porsch-Oezcuermes M., Liebisch G., Kapinsky M., Diederich W.,
RA Dobnik W., Dean M., Allikmets R., Schmitz G.;
RT "ABCG1 (ABC8), the human homolog of the Drosophila white gene, is a
RT regulator of macrophage cholesterol and phospholipid transport.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:817-822(2000).
RN [9]
RP MEDLINE=21474438; PubMed=11590207;
RX MEDLINE=21474438; PubMed=11590207;
RA Schmitz G., Langmann T., Heimerl S.;
RT "Role of ABCG1 and other ABCG family members in lipid metabolism.";
RL J. Lipid Res. 42:1513-1520(2001).
CC -!- FUNCTION: Transporter involved in macrophage lipid homeostasis. Is
CC an active component of the macrophage lipid export complex. Could
CC also be involved in intracellular lipid transport processes. The
CC role in cellular lipid homeostasis may not be limited to
CC macrophages.
CC -!- SUBUNIT: May form heterodimers with several heterologous partners
CC of the ABCG subfamily.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Predominantly
CC localized in the intracellular compartments mainly associated with
CC the endoplasmic reticulum (ER) and Golgi membranes.
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=7;
CC Comment=Additional isoforms seem to exist;
CC Name=1;
CC IsoId=P45844-1; Sequence=Displayed;
CC Name=2; Synonyms=J;
CC IsoId=P45844-2; Sequence=VSP_000047, VSP_000051;
CC Name=3; Synonyms=ABDE;
CC IsoId=P45844-3; Sequence=VSP_000048, VSP_000051;
CC Name=4; Synonyms=G;
CC IsoId=P45844-4; Sequence=VSP_000051;
CC Name=5; Synonyms=F;
CC IsoId=P45844-5; Sequence=VSP_000049, VSP_000051;
CC Name=6; Synonyms=Hi;
CC IsoId=P45844-6; Sequence=VSP_000046, VSP_000051;
CC Name=7; Synonyms=C;
CC IsoId=P45844-7; Sequence=VSP_000050, VSP_000051;
CC -!- TISSUE SPECIFICITY: EXPRESSED IN SEVERAL TISSUES.
CC -!- INDUCTION: Strongly induced in monocyte-derived macrophages during
CC cholesterol influx. Conversely, mRNA and protein expression are
CC suppressed by lipid efflux. Induction is mediated by the liver X
CC receptor/retinoid X receptor (LXR/RXR) pathway.
CC -!- SIMILARITY: Belongs to the ABC transporter family. ABCG (White)
CC subfamily.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; X91249; CAA62631.1; ALT_INIT.

```

```

Query Match      21.1%; Score 706.5; DB 1; Length 678;
Best Local Similarity 28.4%; Pred. No. 5.9e-41;
Matches 194; Conservative 155; Mismatches 251; Indels 83; Gaps 23;

3 SSNVEVFVPSQGNNTGPPATASNDL-----KAPT-----EGAV-LSPHNICYRVKLKSGFLP 54
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
38 SSNMEA-----TETDILNGLHKVVDNNLTIAQRPSLPRAAAVNIETPRDUSYVSPGPMW-- 92
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
55 CRKPYEKETILSNIMTKPG-LNALIGPTGGKSLDLVLAARKDPSGLSDVLIINGAPR 113
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
93 -RKGYKTLKGISGFNSGELVAIMGSPGACKSLTMLAILAGYRE-TGWKGAVLINGLPR 150
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
114 PAN-FKNCISGYVVDVVGTLITVRENTQFSAAKLATTTMTNHEKNERINRVIOELGLDK 172
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
151 DLRCFRKVCSCYTWODMLLPHLTVOEAMVMSAHLKLOB--KDEGRREWKVSLITLALGILLS 208
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```



```
FT TRANSMEM 478 498 POTENTIAL.
SQ SEQUENCE 598 AA; 66906 MW; 9D6414E06898E343 CRC64;
Query Match 20.5%; Score 686.5; DB 1; Length 598;
Best Local Similarity 30.3%; Pred. No. 1.2e-39;
Matches 186; Conservative 124; Mismatches 261; Indels 43; Gaps 14;
QY 61 KEILNININGKPG-LNAILGPTGGGKSSLDVLAARKDPS-GLSGDVLINGARPPAN-- 116
DB 7 KEILNININGKPG-LNAILGPTGGGKSSLDVLAARKDPS-GLSGDVLINGARPPAN-- 116
QY 117 -PKNSGVYQDDVVMGTLTVRENIQFSAALRLATT-MTNKKNERNRIVQIEGLDKVA 174
DB 55 KIRENSAFVQDDVVMGTLTVRENIQFSAALRLATT-MTNKKNERNRIVQIEGLDKVA 174
QY 175 DSKVG-TPIRVSGGGRKRTSIGMELITDPSILFDPTTGLDSSSTANAVLLIKMSK 233
DB 125 DTVIGIPNQLKSGCKKRLSFASEILTCPKILFCDEPTSGLDFAFWAGHVQALRSLAD 184
QY 234 QGRTIIFSHQPRYSIFKLFDSLTLLASRLMFHGPQAEALCYFESAGYHCEAYNPADF 293
DB 185 NGMTVIIHQSSHVSFLNNVCLMAGRVILYLGQDQAVPLEKCGYPCPAYNPADH 244
QY 294 F---LDIINGDSATAVALNREEDFKATEIEPSKQDKPLIEKAEIYVNSPYKTKA--- 347
DB 245 LIITLAVIDSD-----RATSM-----KTIKIROGFLSTDLGQSVLAIGN 284
QY 348 --ELHQLS--GGEKKKTIYFKEISYITSPCHQLRWYSKSFKNLGNPQASIAQIIVT 402
DB 285 ANKLRAAFVTSQDSITSEKTFNFQDYNASFWTQFLAFWRSWLTIVIRDPNLLSVRLLOI 344
QY 403 VVLGVIGAIYFGLKNDSTGIQNRAGVLFLLTNCQFS-SVSAVELFVVEKKLFIHEYIS 461
DB 345 LITAFITGIVFOTPVTPATIIISINGINFNIRNNFMQLQPNVPVITAEPLVIRENAN 404
QY 462 GYRVSVSVGLKSLDPLMTWLPSTIIFCIYVFMGLKPKRADAFVFMVFLMWVAYSAS 521
DB 405 GYRTSAYFLAKNIAELPOYIILP-ILYNTIVYMSGLYFNFWTCFASLVTILITNVAI 463
QY 522 SMALAIAGQSVVSVATLMTICFVFMIFSGLLVNLTTIASWLSWLYQFIPRYGFTAL 581
DB 464 SISYAVATIFANTDVAMTILPIFVVPIMAFGFFITFADIPSYFKWLSVYFKYGEAL 523
QY 582 QHNEFLGQNFCLNATGNNPCNYATCTGEEVLVKGQDLSPWGLKKNHVALACMIVFL 641
DB 524 AINEMDSIKVIECFNNSMTAFALDSCPKNGHQVLESIDFSASHKIFDISILFGMFIGIR 583
QY 642 TIAYKLKFLKXYS 655
DB 584 IIAVYVALL-IRSYN 596
RESULT 14
ID ABG5_HUMAN STANDARD; PRT; 651 AA.
AC Q9H22;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE ATP-binding cassette, sub-family G, member 5 (Sterolin-1).
GN ABCG5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANT GLU-604.
RC TISSUE=Liver;
RX MEDLINE=2053648; PubMed=11099417;
RA Berge K.E., Tian H., Graf G.A., Yu L., Grishin N.V., Schultz J.,
RA Kwitterovich P., Shan B., Barnes R., Hobbs H.H.;
RT "Accumulation of dietary cholesterol in sitosterolemia caused by
RT mutations in adjacent ABC transporters.";
```

```
Science 290:1771-1775 (2000).
[2]
SEQUENCE FROM N.A., VARIANTS SITOSTEROLEMIA HIS-389; HIS-419 AND
PRO-419, AND VARIANT GLU-604.
TISSUE=Liver;
MEDLINE=20578753; PubMed=11138003;
Lee M.-H., Lu K., Hazard S., Yu H., Shulien S., Hidaka H., Kojima H.,
Allkmetes R., Sakuma N., Pegoraro R., Srivastava A.K., Salen G.,
Dean M., Patel S.B.;
"Identification of a gene, ABCG5, important in the regulation of
dietary cholesterol absorption.";
Nat. Genet. 27:79-83 (2001).
[3]
REVIEW.
MEDLINE=2147438; PubMed=11590207;
Schmitz G., Langmann T., Heimerl S.;
"Role of ABCG1 and other ABCG family members in lipid metabolism.";
J. Lipid Res. 42:1513-1520 (2001).
[4]
VARIANTS SITOSTEROLEMIA GLN-146; HIS-389; PRO-419; HIS-419 AND
SER-550, AND VARIANT GLU-604.
MEDLINE=21344600; PubMed=11452359;
Lu K., Lee M.-H., Hazard S., Brooks-Wilson A., Hidaka H., Kojima H.,
Oee L., Stalenhoef A.P.H., Miettinen T., Bjorkhem I., Bruckert E.,
Pandya A., Brewer H.B. Jr., Salen G., Dean M., Srivastava A.K.,
Patel S.B.;
"Two genes that map to the STSL locus cause sitosterolemia: genomic
structure and spectrum of mutations involving sterolin-1 and
sterolin-2, encoded by ABCG5 and ABCG8, respectively.";
Am. J. Hum. Genet. 69:278-290 (2001).
-!- FUNCTION: transporter that appears to play an indispensable role
in the selective transport of the dietary cholesterol in and out
of the enterocytes and in the selective sterol excretion by the
liver into bile.
-!- SUBUNIT: May form heterodimers with ABCG8 or be tightly coupled to
ABCG8 along a pathway regulating dietary-sterol absorption and
excretion.
-!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
-!- TISSUE SPECIFICITY: Strongly expressed in the liver, lower levels
in the small intestine and colon.
-!- DISEASE: Defects in ABCG5 are a cause of sitosterolemia
[IMM:210250]; also known as phytosterolemia or shellfish
sterolemia. It is a rare autosomal recessive disorder
characterized by increased intestinal absorption of all sterols
including cholesterol, plant and shellfish sterols, and decreased
biliary excretion of dietary sterols into bile. Sitosterolemia
patients have hypercholesterolemia, very high levels of plant
sterols in the plasma, and frequently develop tendon and tuberos
xanthomas, accelerated atherosclerosis and premature coronary
artery disease.
-!- SIMILARITY: Belongs to the ABC transporter family. ABCG (White)
subfamily.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
```

EMBL; AF320293; AAC40003.1; -;  
EMBL; AF312715; AAC53099.1; -;  
Genew; HGNC:13886; ABCG5.

MIM; 605459; -;  
MIM; 210250; -;  
GO; GO:0030299; P.cholesterol absorption; NAS.

InterPro; IPR003593; AAA\_Arase.

InterPro; IPR003439; ABC\_Transporter.

ProDom; PD000005; ABC\_tran; 1.

ProDom; PD000006; ABC\_transporter; 1.

SMART; SM00382; AAA; 1.

PROSITE; PS00211; ABC\_TRANSPORTER\_1; FALSE\_NEG.



DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 1.  
 KW ATP-binding; Glycoprotein; Transmembrane; Transport; Polymorphism;  
 KW Disease mutation.  
 FT DOMAIN 1 383 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 384 404 1 (POTENTIAL).  
 FT DOMAIN 405 421 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 422 442 2 (POTENTIAL).  
 FT DOMAIN 443 462 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 463 483 3 (POTENTIAL).  
 FT DOMAIN 484 503 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 504 524 4 (POTENTIAL).  
 FT DOMAIN 525 548 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 549 569 5 (POTENTIAL).  
 FT DOMAIN 570 591 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 592 612 6 (POTENTIAL).  
 FT DOMAIN 613 633 CYTOPLASMIC (POTENTIAL).  
 FT NP\_BIND 634 651 ATP (POTENTIAL).  
 FT CARBOHYD 652 684 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 685 718 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARIANT 719 746 E -> Q (in sitosterolemia).  
 FT VARIANT 747 774 /FTID=VAR\_012244.  
 FT VARIANT 775 802 R -> H (in sitosterolemia).  
 FT VARIANT 803 830 /FTID=VAR\_012245.  
 FT VARIANT 831 858 R -> H (in sitosterolemia).  
 FT VARIANT 859 886 /FTID=VAR\_012246.  
 FT VARIANT 887 914 R -> P (in sitosterolemia).  
 FT VARIANT 915 942 /FTID=VAR\_012247.  
 FT VARIANT 943 970 R -> S (in sitosterolemia).  
 FT VARIANT 971 998 /FTID=VAR\_012248.  
 FT VARIANT 999 1026 Q -> E.  
 FT VARIANT 1027 1054 /FTID=VAR\_012249.  
 FT SEQUENCE 651 AA; 72503 MW; 950BABCBB6A1536 CRC64;  
 Query Match 20.4%; Score 682.5; DB 1; Length 651;  
 Best Local Similarity 29.2%; Pred. No. 2.5e-39;  
 Matches 182; Conservative 138; Mismatches 249; Indels 55; Gaps 18;  
 QY 13 SQNTNGFPATANDLKAFTGAVLSFHNICRVKLSGFLPCRKPEVEKILSNINGIMK 72  
 DB 21 SQSLGAPATAP---EPHSLGLHASYSVSHRVPWMDITSCQOWTRQILKDSLYVE 77  
 QY 73 PG-LNAILGPTGGKSLDLVAARDPSG-LSGDVLING-APRPANFKCNGYVVQDDV 129  
 DB 78 SGQIMCILSGSGKTTLLDAMSGRLGRAGTFLGEVYVNGRALRRQFQDCFSYVLQSDT 137  
 QY 130 VMGTLVRENLPQSAALRLATTWNHEKNERINRVQELGLKVDKSVGTQFIRGSGG 189  
 DB 138 LLSLTVRETHYTLALLAIRGNPG-SPOKKVEAVNAE-SLSHVADRLIGNYSLOGISTG 196  
 QY 190 ERKRTSIGMELITDPSLFLDEPTTGLDSTANAVLLKRMKSGQRTIIFGTHOPRYSI 249  
 DB 197 ERRVSIAGLLQDPKVMLEDETTGLDCMTANQIVVLLVLAARNRIVVLIHQPSREL 256  
 QY 250 FKLFDLSLTLASGRLMFHPQAQALGYFSGAGVHCAYNNPADFFDIIDNGSTAVALNR 309  
 DB 257 FQLFDKIALISFGLIFCGTPAEMLDFFNDGCGYPCPSHNSPDFYMDLTSVDTQ----SK 312  
 QY 310 EEDFKATEIIEPSKODKFLKELAEIYNSSFYKFKAEHLQSGGKKKKTIVPKESY 369  
 DB 313 ERE-----IETSKR-----VOMTESAYKKAICHKT-----LKNIERMKHKLTPDMVFP 356  
 QY 370 TT----SFCQLRWKSRSPKLNGLNQPQASIAQIIVTVVLGLVIGAIYFGLKNDST---- 421  
 DB 357 KTKDSPGVFSKGLGVLLRVRTRNLVRNKLAVITRLQLNGLFL--LFFVLVRSNVLKG 414  
 QY 422 GIGNRAGVLP-FLTTWQCFSSVAVELFVVEKFLFTHIYSIGYRVSSYFGLKLSLDLAP 480  
 DB 415 AIQDRVGLLYQFVGATPYTGMLNAVNLFPVLRVAVSDQESODGLYQKQMLVAL--HVLIP 473  
 QY 481 MTLVLSIIFTCYIVFVGLPKPKADAFVMMFTLM--MVAYSASSMALAAAGQSVSVVA 537  
 DB 474 FSVVATMIFSSVCVWTLLGLHPEVARFGYFSAALLAPHLIGFELTLLGLGIVQNPENIVNSV 533

QY 538 TLLMTICFVFMWLFSGLLVNLTTIASWLSLQVIESIPRYGVTALQHNEFLGNFCPG--- 594  
 DB 534 VALLSIAGV--LVGSGFLRNQEMPIPKLIISFTFKICSEILVNVNFGVGLNFTGSSN 591  
 QY 595 LNATGNPNPCNYA-----TCTG 610  
 DB 592 VSVTNPMACTQIGIQIENKTCPG 615  
 RESULT 15  
 'ABGS\_RAT  
 ID ABGS\_RAT STANDARD; PRT; 652 AA.  
 AC Q99PE7; Q8C1Q4;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE ATP-binding cassette, sub-family G, member 5 (Sterolin-1).  
 GN ABCG5.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley; TISSUE=Small intestine;  
 RX MEDLINE=20578753; PubMed=11138003;  
 RA Lee M.-H., Lu K., Hazard S., Yu H., Shulenin S., Hidaka H., Kojima H.,  
 RA Allikmets R., Sakuma N., Pegoraro R., Srivastava A.K., Salen G.,  
 RA Dean M., Patel S.B.;  
 RA "Identification of a gene, ABCG5, important in the regulation of  
 RT dietary cholesterol absorption.";  
 RL Nat. Genet. 27:79-83 (2001).  
 RN [2]  
 RP REVISION TO 2.  
 RA Lu K., Lee M.-H., Patel S.B.;  
 RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A. TISSUE SPECIFICITY, AND VARIANT CYS-583.  
 RC STRAIN=GH, SHR, SHRSP, Sprague-Dawley, Wistar, Wistar Kyoto, and WKA;  
 RX PubMed=12783625;  
 RA Yu H., Pandit B., Klett E., Lee M.H., Lu K., Helou K., Ikeda I.,  
 RA Egashira N., Sato M., Klein R., Batta A., Salen G., Patel S.B.;  
 RT "The rat STSL locus: characterization, chromosomal assignment, and  
 RT genetic variations in sitosterolemic hypertensive rats.";  
 EL BMC Cardiovasc. Disord 3:4-4 (2003)  
 CC -!- FUNCTION: Transporter that appears to play an indispensable role  
 CC of the selective transport of the dietary cholesterol in and out  
 CC of the enterocytes and in the selective sterol excretion by the  
 CC liver into bile.  
 CC -!- SUBUNIT: May form heterodimers with ABCG8 or be tightly coupled to  
 CC ABCG8 along a pathway regulating dietary-sterol absorption and  
 CC excretion (by similarity).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
 CC -!- TISSUE SPECIFICITY: Expressed only in liver and intestine.  
 CC -!- POLYMORPHISM: The polymorphism at position 583 is found in strains  
 CC SHR, SHRSP and Wistar Kyoto which are both hypertensive and  
 CC sitosterolemic. Strains which are hypertensive but not  
 CC sitosterolemic do not contain a polymorphism at this position.  
 CC -!- SIMILARITY: Belongs to the ABC transporter family. ABCG (White)  
 CC subfamily.  
 ---  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 ---  
 CC EMBL; AF312714; AAG5098.3; -;  
 CC EMBL; AY145899; AAN64275.1; -;  
 CC InterPro; IPR003593; AAA\_ATPase.  
 CC InterPro; IPR003439; ABC\_transporter.

Search completed: March 23, 2004, 18:30:03  
Job time : 20 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

QM protein - protein search, using sw model  
Run on: March 23, 2004, 18:27:43 ; Search time 20 Seconds  
(without alignments)  
3150.272 Million cell updates/sec

Title: US-09-961-086-1  
Perfect score: 3352  
Sequence: 1 MSSSNVEVFPVSGQNTNGF.....MIVFLTIAYLKLLFLKKYS 655

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues  
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78.\*  
1: PIR1.\*  
2: PIR2.\*  
3: PIR3.\*  
4: PIR4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2849.5	85.0	656	JC7860	brain multidrug re
2	835.5	24.9	1049	S19421	ATP-dependent perm
3	812	24.2	687	FFFW	white protein - fr
4	800.5	23.9	737	T46101	ABC transporter-11
5	774	23.1	646	C86441	probable ABC trans
6	767.5	22.9	687	D96553	hypothetical prote
7	746	22.3	649	A84509	probable ABC trans
8	745	22.2	725	T47652	ABC transporter-11
9	741.5	22.1	739	T45891	ABC transporter-11
10	739	22.0	678	H96552	hypothetical prote
11	726.5	21.7	708	T47650	ABC transporter-11
12	721	21.5	635	T08934	hypothetical prote
13	720.5	21.5	755	C84791	probable ABC trans
14	717.5	21.4	740	T02567	probable ATP-bindi
15	712.5	21.3	609	E96742	probable ABC trans
16	710	21.2	638	G02068	white homolog - hu
17	708	21.1	646	TJ7777	ATP binding casset
18	707.5	21.1	547	TJ1543	hypothetical prote
19	707.5	21.1	720	T47648	ABC transporter-11
20	677	20.2	559	B98474	protein C05D10.3 [
21	676	20.2	725	C84423	probable ABC trans
22	669	20.0	659	E96313	hypothetical prote
23	664	19.8	608	T34391	hypothetical prote
24	660	19.7	662	T47649	ABC transporter-11
25	658.5	19.6	1294	T7690	probable membrane
26	657.5	19.6	1450	T45898	ABC transporter-11
27	656.5	19.6	590	T24588	protein F12M16.17
28	649.5	19.4	633	T19189	hypothetical prote
29	645.5	19.3	577	T04229	ABC-type transport

30	639	19.1	658	2	T31958	hypothetical prote
31	636.5	19.0	639	2	G88839	protein C10C6.5 [i
32	636.5	19.0	695	2	T21109	hypothetical prote
33	627.5	18.7	610	2	T19333	hypothetical prote
34	612.5	18.3	1501	2	S50992	SNQ2 protein - yea
35	612	18.3	1530	2	S52239	brefeldin A resist
36	603	18.0	1530	2	T52010	hypothetical prote
37	602	18.0	1511	2	A53151	pleiotropic drug r
38	594	17.7	1564	2	S55517	probable transport
39	593	17.7	1443	2	T02431	probable ABC trans
40	590	17.6	1469	2	H96622	probable ABC trans
41	585	17.5	1333	2	S63403	probable membrane
42	580.5	17.3	1420	2	T02644	ABC-type transport
43	580.5	17.3	1529	2	S69688	hypothetical prote
44	572	17.1	1413	2	G84790	probable ABC trans
45	564	16.8	1466	2	T30566	ATP-binding casset

ALIGNMENTS

RESULT 1

JC7860  
brain multidrug resistance protein, BMDP - pig  
C:Species: Sus scrofa domestica (domestic pig)  
C:Date: 18-Nov-2002 #sequence\_revision 18-Nov-2002 #text\_change 31-Mar-2003  
C:Accession: JC7860  
R:Eisenblatter, T.; Galla, H.J.  
Biochem. Biophys. Res. Commun. 293, 1273-1278, 2002  
A:Title: A new multidrug resistance protein at the blood-brain barrier.  
A:Reference number: JC7860; MUID:22050127; PMID:12054514  
A:Accession: JC7860  
A:Molecule type: mRNA  
A:Residues: 1-656 <EIS>  
A:Cross-references: GB:AJ420927  
A:Experimental source: brain  
C:Comment: This protein, a new transport protein of the ATP-binding cassette (ABC) sup  
exclusion of xenobiotics from the brain and participates in drug transport across the  
C:Genetics:  
A:Gene: bmdp

Query Match	85.0%	Score	2849.5	DB 2	Length	656			
Best Local Similarity	84.3%	Pred. No.	4.9e-195						
Matches	553	Conservative	44	Mismatches	58	Indels	1	Gaps	1
QY	1	MSSSNVEVFPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICYRVKLSGFLPCRPEVE	60						
Db	1	MSSSNVQVSIPLMSKNTNGLPGSSSNEILKTSAGGAVLSFHDICIRVKVSGFLFCRKIVE	60						
QY	61	KEILSNINGIMKPGNLAILGPTGGKSSLLDVLAAKDPGSLGSDVLINGAPRANFKCN	120						
Db	61	KEILTNINGIMKPGNLAILGPTGGKSSLLDVLAAKDPHGLSGDVLINGAPRANFKCN	120						
QY	121	SGYVQDDVVMGTLTVRENLOFSAALRLATMTWHEKNERINRVIOELGLDKVADSKVGT	180						
Db	121	SGYVQDDVVMGTLTVRENLOFSAALRLPTMTWHEKNERINRVIOELGLDKVADSKVGT	180						
QY	181	QPIRGVSGGERKRTSIGMELITDPSILFDEPTTGLDSTANAVLLILKRMKSQKQRTIIF	240						
Db	181	QPIRGVSGGERKRTSIAMELITDPSILFDEPTTGLDSTANAVLLILKRMKSQKQRTIIF	240						
QY	241	SIHQPRYSIFKLFDSLTLLAGSLMFHGPAPQALGYFESAGYHCEAYNNPADDFFLDING	300						
Db	241	SIHQPRYSIFKLFDSLTLLAGSLMFHGPAREALGYFASIGYCNCPYNNPADDFFLDVING	300						
QY	301	DSTAVALNR-EDDFKATIEIPEKDKPLIEKLEIYVNSSSFYKETAELHQLSGGEKKK	359						
Db	301	DSSAVULSRADRDGAQEPSEPEKDTPLIDKLAIFYNSSFDDTKVELQDFSGGRKKK	360						
QY	360	KITVFKGISYTTFSCHQLRWVSRKRSFKNLLGNPQASIAQIITVTVLGLVIGAIYFGLKND	419						
Db	361	KSSVYKEVYTTFSCHQLRWISRRSFKNLLGNPQASVAQIITVITILGLVIGAIYFDLKND	420						

355 LGSSKSPRLP-DEDAVNNFLQNEDDTL-----ATLSPENITYSPVINS-----DGVE 401

61 KEILSNINGIMKPG-LNAILOPTGGGKSSLLDLVLAARKDPSGLSDVLINGARPP-ANFK 118

403 ETVLNEISGIVKPGQILAIIMGGGAGATKTLILDILAMKRTGHHVSGISKVGINGISMDRKSFS 462

119 CNSGVVQDDVVMGTLTVRENLOFSAALRLATTMTWHEKNERNRVIQELGLDKVADSKV 178

463 KIIGFVQDDFLLPLTAVFETVLSALLRLPKALSFEAKRKYVYLEELRIIDIKRII 522

179 GTOFTRGVSGGERKTSIGMELITDPSILFDEPTTGLSDSTANAVLLLLKMSKO-GRT 237

523 GNEFDRGISGGEKRVSIACELVSPVLFLDEPTSGLDASNANNVIECLVRUSSYNTK 582

238 IFTSIHQPRYSIFKFLDPSLTLASGRMLFHGPAQEAALGYEPESAGYHCEAYNNPADPFLDI 297

583 LVLSIHQPRSNIFYFLFDKLVLLSGEMVYSGNAKVSFEFLRNEGYSICPDNYNIADVLDI 642

298 -----INGDPSTAV 305

643 TTEAGPQKRRIRNISDLEAGTNDIDNTIHOHTTTSSDGTTOREWAHLAHRDEIRS 702

306 ALNREEDFKATE-----IIEPSQDKPLIEKLAETIYVNSFYKETAFLHQ-LSGGEKKKK 360

703 LLRDEEDVEGTDGRRGATEIDLNTKLLHDK-----YKDSVYVYAEILSQIEEVLSEGDEESN 758

361 IT--VPEKISYTTSECHOLRWVSKRSFKULLGNFQASIAQIIVTVVLGVIGAIYFLGN 418

759 VINGDLPTGQSGAGFLQSLINSRSFKMTRNPKLLGNLYLTLILSLFEGILYINVSN 818

419 DSTGQNAGVLPFLTTNQCFSSVSAVELFVVEKKLFIHEYISGYRYRVSSYFLGKLSDL 478

819 DISGQNRMLGFFFTLVTFGVTGLSSPALERIIIFKERSNNYSPLAYY-SKIMSEV 878

479 LPMTMLPSIIFTCIYVFMGLKPKADAFVVMFPLVMVAYSASSMALAIAGOSVUSVAT 538

879 VPLRVPPILLUSLIVPMTGNMKDNFAKCGILGLIILFNIGLSLEILITIGIIFEDLNNSI 938

539 LMLTICFVPMIFSGLLV---NLTTIASLWSLQYFSIPRYGTALQHNEF----- 586

939 ILSVLVLGSLFLSGLFINTKNTINVA--PKYLNKFSVFYAYVESLLINEVKTMLKERE 996

587 LQGNP-CPGLNATGNPCNYATCTGEEVLVKQGI--DLSPWGLWKHNVALACMIVFLTI 643

997 YGLNIEVPG-----ATILSTGFGVVQNIIVFDIK-----ILALFNVPFLIM 103

644 AYKLKLLFL 651

1037 GYLALKWI 1044

RESULT 3

FYFW

white protein - fruit fly (*Drosophila melanogaster*)

C:Species: *Drosophila melanogaster*

C:Date: 31-Dec-1990 #sequence revision 17-Feb-1995 #text\_change 19-Jan-2001

C:Accession: S08635; S07263; S10240

R:Pepling, M.; Mount, S.M.

Nucleic Acids Res. 18, 1633, 1990

A:Title: Sequence of a cDNA from the *Drosophila melanogaster* white gene.

A:Reference number: S08635; MUID:90221897; PMID:2109311

A:Accession: S08635

A:Molecule type: mRNA

A:Residues: 1-687 <REP>

A:Cross-references: EMBL:X51749; NID:g8825; PIDN:CAA36038.1; PID:g8826

R:O'Hare, K.; Murphy, C.; Lewis, R.; Rubin, G.M.

J. Mol. Biol. 180, 437-455, 1984

A:Title: DNA sequence of the white locus of *Drosophila melanogaster*.

A:Reference number: S07263; MUID:85134865; PMID:6084717

A:Accession: S07263

A:Molecule type: DNA

A:Residues: 1-24, 11FEIPEYHCRVTAD', 30-334, 'ITLHNSIPAWPESVLPTTIRKFTTYRCWPLCPDGEDG

A:Cross-references: EMBL:X02974

A:Experimental source: strain Canton S  
 R:O'Hare, K.  
 Submitted to the EMBL Data Library, June 1985  
 A:Reference number: S10240  
 A:Accession: S10240  
 A:Molecule type: DNA  
 A:Residues: 1-24, 'LIFEIPVHCVTAD', 30-687 <OHA2>  
 A:Cross-references: EMBL:X02974, NID:g10873, PIDN:CAA26716.1, PID:g10874  
 A:Experimental source: strain Canton S  
 C:Genetics:

A:Gene: white; w  
 A:Cross-references: FlyBase:FBgn0003996  
 A:Introns: 24/3; 116/1; 334/2; 439/3; 483/3  
 C:Superfamily: fruit fly white protein; ATP-binding cassette homology  
 C:Keywords: Atp; glycoprotein; nucleotide binding; P-loop; transmembrane protein  
 F:113-317/Domain: ATP-binding cassette homology <ABC>  
 F:130-137/Region: nucleotide-binding motif A (P-loop)  
 F:261-265/Region: nucleotide-binding motif B  
 F:67,93,472,554,651/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 24.2%; Score 812; DB 1; Length 687;  
 Best Local Similarity 32.1%; Pred. No. 7.7e-50;  
 Matches 210; Conservative 134; Mismatches 251; Indels 60; Gaps 17;  
 74 NMDIFGAVNQ-----PQSGWRQLVNRTRGLFCNERHI-----PAPR---KHL 113  
 65 SNINGIMKPG-LNALIGPTGGKSSLDVLAARKDPSGL-----SGDVINGAPRPA-NFX 118  
 114 KNCVGAVPGELLAVMGSSGAGKTTLLNALAFR-SPOGIQVSPGMRLLNGQPVDAKEMQ 172  
 119 CNSGYVQDDVVMGTLVRENLOFSAALRLATTNNEKNERINRVIOELGLDKVADSK 178  
 173 ARCAVQDDLFGLSLTAREHLIFQAVMRPHLTYPQVRVARVDVIOELSLSKCQHTII 232  
 179 GTQ-FIRGVSQGGKRTSIGNELTDPISILFDEPTTGLDSSANAVILLIKRMSKQGT 237  
 233 GVPGRVGLSGGERKRLAFASEALTDPLTICDEPTSDFTSAHSVQVLKLSQKGT 292  
 238 IIFSIHQPRYSIFKLFDSLTLASGRLMFHGPAQALGYFESAGYHCEAYNNPADFLDI 297  
 293 VILTHQPSSELFELFDKILLMAGRVAFLGTPSEAVDFSYVGAQCTNYPNPAFVQV 352  
 298 INGOSTAVALNREDFKATEIIEPSKQDKPLIEKLIAYNNSPYKETKABLHQLSGEK 357  
 353 L-----AVVPGREIESR-----DRIAKICDNFAISKVAR-DMEQLLATKN 391  
 358 KKKITVPKEISYT--TSFCHQLRWVSKRSFKNLLGNPOASTAQIIVTWLGLVIGATYFG 415  
 392 LEKPLEQENGYTYKATWFMQFRAVLWRSLWSVLKEPLLKVRLIQTWVAIIIGLIFLG 451  
 416 LKNDSGTIONRAGLVFFLTNNQCFSSVA-VELFVVEKKLFIHEYISGYVRVSSYFLGKL 474  
 452 QQLTQGVWNINGAIFLFTNMTQNVFATINVTSELPLVPMREARSLRYCDDTYFLGKT 511  
 475 LSDLLPMTLSIIPTCIIVYFMGLKPKADAFFVMMFTLMVAYSASSMALAAGOSVV 534  
 512 IAE-LPLFLTVLFTAYPMIGLRAGLVHFFNCLALVILVANVSTSGFLVISCASSST 570  
 535 SVATLLMTICFVFMIFSGLLVNLTTIASWLSWLYQFSIPRYGFTALQHNFLGQNPCPG 594  
 571 SWALSGVPPVPIPFLLFGGFLNSGVPVYLKWSLWSFRYANEGLLINQWADVE--PG 628  
 595 -LNATGNPCNVATCTGEYLVKQIDLSPLGKLNKHNVALACMIVILFTAYLKL 648  
 629 EISCTSSN-----TTCFSSGKKVILETLNFSAADPLDLYVGLAILIVSPRVLAYLAL 679

RESULT 4

T46101  
 ABC transporter-like protein - Arabidopsis thaliana  
 N:Alternate names: protein T25B15.80

C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Feb-2000  
 C:Accession: T46101  
 R:Alcaraz, J.P.; Clabault, G.; Cottet, A.; Mache, R.; Mewes, H.W.; Lemcke, K.; Mayer,  
 submitted to the Protein Sequence Database, January 2000  
 A:Reference number: Z23021  
 A:Accession: T46101  
 A:Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-737 <ALC>  
 A:Cross-references: EMBL:AL132972  
 A:Experimental source: cultivar Columbia; BAC clone T25B15  
 C:Genetics:  
 A:Map position: 3  
 A:Introns: 122/1; 146/3; 225/2; 277/2; 338/3; 422/2; 535/1; 628/3; 664/3  
 A:Note: T25B15.80

Query Match 23.9%; Score 800.5; DB 2; Length 737;  
 Best Local Similarity 31.4%; Pred. No. 5.6e-49;  
 Matches 211; Conservative 132; Mismatches 227; Indels 101; Gaps 21;  
 27 DLKAFTEGAV-----LSFHNICYRVKLKSGFLPCRKPKVEKEILSININGIMKPG- 74  
 124 DIERATSSVVFQAEPTFPIYLKIDITYKVTYG-----WTSSEKSLNIGISSAYGE 179  
 75 LNALIGPTGGKSSLDVLAARKDPSGLSDVINGAPRPAFNCNSGYVQDDVVMGTL 134  
 180 LIALMGPSGSGKTTLLNALGGRFNQONIGSGSVYNDKPKYKHLKTRIGFTVQDDVLP 239  
 135 TVRENLOFSAALRLATTNNEKNERINRVIOELGLDKVADSKYGTQFIRGVSQGERKRT 194  
 240 TVKETLTALLRLPFTLITEQEKQRAASVIOELGLERCQDTMIGGSFVRGVSQGERKRV 299  
 195 SIGMELITDPSILFDEPTTGLDSSANAVILLIKRMSKQGTIIFSIHQPRYSIFKLPD 254  
 300 CIGNEIMTNPSLLLDDEPTSLDSTALKIVQMLHCAKAGTIVTTHQPSRLFHFPD 359  
 255 SUTLASGRLMFHGPAQALGYFESAGYHCEAYNNPADFLDIINGDSTAVALNREDFK 314  
 360 KLWLSRGLSLYFGKASEAMSYFSSIGCSPLLAMPNPAFLDLVNGNNDIS----- 411  
 315 ATEIIEPSKQDKPLIEKLIAYNNS-----SPYKETKABLHQLS-----GGE 356  
 412 ----VPSALKERKMKIIRL-ELYVRNVKCDVETQYLEEAYKTQIAVWERMKLMAPVLD 466  
 357 KKKKITVPKE-----ISYTSFCHQLRWVSKRS-----FKNLLGNPOASTAQIIVTW 409  
 467 VKLMTITCPKREWGLSWEQYCLSLRGIKERRHDYFSWL-----RVTQVLSTAI--- 517  
 410 GAIYFGLKNDSTGIO-NRAGLVFFLTNNQCFSSV-SAVELFVVEKKLFIHEYISGYVRV 467  
 518 GLLWN--QSDITTSQRTSRGSLFFIAVFWGFFPVFTAFTPQERAMLSKERESNNYRLS 575  
 468 SYFLGKLSDLLPMTLSIIPTCIIVYFMGLKPKADAFFVMMFTLMVAYSASSMALAI 527  
 576 AYFVARTTSD-LPLDLILPVLVYVYFMAGLRSAESFFLSVLTVFCIVAAQGLGLAI 634  
 528 AAGQSVSVATLLMTICFVFMIFSGLLVNLTTIASWLSWLYQFSIPRYGFTAL---QHN 584  
 635 GASLMDLKKATTLASTVNTFMTLAGGYFVK--KVFFIAWIRFMSFNHTYKLLVKKVQE 692  
 585 EFLGQNFPCPLNATGNPCNVATCTGEYLVKQIDLSPLGKLNKHNVALACMIVILFTIA 644  
 693 EIM-----ESVNGEE--IBSGL-----KEVSALVAMIIGYRLVA 724  
 645 YLKLFLPKKYS 655  
 725 YFSLRMRMKLHS 735

RESULT 5

C86441  
 Probable ABC transporter [imported] - Arabidopsis thaliana













Qy 358 -----KKKIVFKEISYTTTFCHQLRWVSKRSFKNLLGNPOASIAQIIIVTVVLGIVGA 411  
Db 411 VAHGGAATNTTTLAVAPANPMWIEIKTLKSGMNSRQPELPGIRIASVUITGILAT 470  
Qy 412 IYGLKNDSTGIONRAGVLFLLTNOCFSSVSAVELFVVEKKLFIHEYISGYRVSSYFL 471  
Db 471 VFWRLNSPKGVQERLGFAPAFAMSTWFTYTCADALPVFLQERYIFRMETAYNAYRSSYVL 530  
Qy 472 GKLLSDLLPMTLPSIIFTCIVFVLMGLKPKADAFVMMFTLMMVAYSASSMALAIAAQ 531  
Db 531 SHAIVS-PPSLIFLSVAFATTTWAGLDGLTGLFYCLIIILASFWSSGFVFLSGVV 589  
Qy 532 SVVSATATLMTICFVFMVMPGSLVNLTTIASWMLQVPSIPRYGFTALQHNFLGQNF 591  
Db 590 PSVMGLGYTIVAILAYFLFSGFFINRIPDIWIFHYMSLVKPYEAVLQNEFSDATK 649  
Qy 592 C--PGLNATGNP-----CNVATC--TGEVYLKQG-IDLSP 623  
Db 650 CFVRGVQIFNTPLGELPEVMKLLGTGVSXSLGVTSSTTCLTQSDILRQGVQLSK 709  
Qy 624 WGLMKNHVALCMIVIFLTIA---YLKLF 650  
Db 710 WN-----CLFITVAFGFFRILF 727

RESULT 15  
E96742  
probable ABC transporter F17M19.11 [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001  
C;Accession: E96742  
R;Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A;Reference number: A86141; MUID:21016719; PMID:11130712  
A;Accession: E96742  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-609 <STC>  
A;Cross-references: GB:AE005173; NID:g6978921; PIDN:AAF34313.1; GSPDB:GN00141  
C;Genetics:  
A;Gene: F17M19.11  
A;Map position: 1  
C;Superfamily: fruit fly white protein; ATP-binding cassette homology

Query Match 21.3%; Score 712.5; DB 2; Length 609;  
Best Local Similarity 31.0%; Pred No 8e-43;  
Matches 190; Conservative 125; Mismatches 256; Indels 41; Gaps 16;

Qy 56 RKPVEKILNINIMKPG-LNALGPTGGKSLDLVLAARKDPSGLSDVLINGAPRP 114  
Db 23 RSTEERTILSGVTGMISPGEFMAVLGPSSGSKSTLLNAVAGRLHGSNLTKILINDGKIT 82  
Qy 115 ANPKCSGYVDDVVMGTLTVRENLQPSAALRLATWTNHEKNERINRVIQELGLDKVA 174  
Db 83 KQTLKRTGFGVAQDQLLPHLTVRETLVFLVALLRPLRLTRDVKLRAESVISBELGTRCE 142  
Qy 175 DSKVGTQFIRGVGGERKRTSIGMELITDPSILFDEPTTGLDSTANAVLLLLKEMSK- 233  
Db 143 NTVVGNFTIRGISGGERKRVSIHAELLINPSLLVLDPTSGLDATAALRLVQTLAGLAHG 202  
Qy 234 QGRTIIPSIHQPRYSIFKLFDSLTLLASGRMLPHGPAQALGYFESAGYHCEAYNPNADF 293  
Db 203 KGKTVVTSIHQPSRSRVQMFDTVLLISEGKCLFVGKGRDAMAYFESVGFSPAPFNNPADF 262  
Qy 294 FLDIING--DSTAVALNREEDFKATEIIEPSKQDKPLIEKLABI-----YVNSFYKETKA 347

Db 263 LLDLANGVQOTDGVTEREKENVRQTLVTAYDITLLAPQVKTCTIEVSHFPQDNARFVKT--- 319  
Qy 348 ELHQLSGGEKKKKTIVFKEISYTTSPCHQL-RWVSKESFKNLLGNPOASIAQIIIVTVVLG 406  
Db 320 ---RVNGGGITTCA-----TWFSQCLILHRLKKERRHESF-----DLRIFQVVAAS 365  
Qy 407 LVIGAIYFGLKNDSTGIONRAGVLFLLTN-QCFSSVSVAVELFVVEKKLFIHEYISGYR 465  
Db 366 ILCGLMMW--HSDYRDVHDLGLLFFISIFWGLPFSFNAVFTFPQERAIPTRRASGMYT 423  
Qy 466 VSSYFLGKLLSDLLPMTLPSIIFTCIVFVLMGLKPKADAFVMMFTLMMVAYSASSMAL 525  
Db 424 LSSYFMARVLGSLSMELVLPASFLT-FTYMWVYLRPGIVFPFLTSLVLLLYVLASQGLGL 482  
Qy 526 AIAAGQSQSVSVATLLMTICFVFMVMPGSLVNLTTIASWMLQVPSIPRYGFTALQHNFLGQNF 582  
Db 483 ALGAAIMDAKKASTIVITVMTLAFVLGTGYVN--KVFSGMVMWKYVSTTFYCYELLVAIQ 540  
Qy 593 HNEFLGQNFPC--GLNATGNPNPCYATCTGEEVLVKQIDLSPLWGLMKNHVALCMIVIF 640  
Db 541 YGS--GEEILRMLGCDKSGKQASAAATSAGCRFVEEVI--GDVGMWTSVGLVFLMPFGY 596  
Qy 641 LTIAYLKILFLK 652  
Db 597 RVLAYLALRRIK 608

Search completed: March 23, 2004, 18:31:36  
Job time : 23 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.  
CM protein - protein search, using sw model  
Run on: March 23, 2004, 18:31:13 ; Search time 55 Seconds  
(without alignments)  
3083.923 Million cell updates/sec  
Title: US-09-961-086-1  
Perfect score: 3352  
Sequence: 1 MSSNVVEFIPVSGQNTNGF.....MIVFIETIAYLKLFLKKYS 655  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
Searched: 1049977 seqs, 258955339 residues  
Total number of hits satisfying chosen parameters: 1049977  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :				Published Applications AA.*			
1:	/cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*	2:	/cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*	3:	/cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*	4:	/cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5:	/cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*	6:	/cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*	7:	/cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*	8:	/cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9:	/cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*	10:	/cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*	11:	/cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*	12:	/cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13:	/cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*	14:	/cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*	15:	/cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*	16:	/cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17:	/cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*	18:	/cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*				

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3352	100.0	655	10	US-09-961-086-1
2	3352	100.0	655	15	US-10-405-806-13
3	3346	99.8	655	9	US-09-981-353-35
4	3346	99.8	655	14	US-10-120-687-61
5	3346	99.8	655	15	US-10-405-806-2
6	3338	99.6	655	9	US-09-866-866A-27
7	3331	99.4	655	14	US-09-866-866A-10
8	3331	99.4	655	14	US-10-090-455-5
9	3053.5	91.1	604	9	US-09-745-763-197
10	2757	82.2	657	9	US-09-866-866A-14
11	835.5	24.9	1049	15	US-10-369-493-1520
12	812	24.2	663	13	US-10-108-605-245
13	805	24.0	1084	12	US-10-424-593-242078
14	805	24.0	1101	12	US-10-425-114-63125
15:	795.5	23.7	676	15	US-10-369-493-3799

16	773.5	23.1	695	12	US-10-424-593-176182
17	746.5	22.3	819	12	US-10-425-114-54421
18	711	21.2	846	14	US-10-090-455-13
19	710.5	21.2	874	14	US-10-090-455-4
20	710.5	21.2	674	16	US-10-429-160-10
21	710	21.2	638	13	US-10-072-621-10
22	710	21.2	646	13	US-10-072-621-9
23	710	21.2	646	14	US-10-090-455-2
24	708	21.1	627	14	US-10-090-455-8
25	708	21.1	646	13	US-10-154-452-4
26	707	21.1	725	12	US-10-424-593-175941
27	700	20.9	646	13	US-10-154-452-8
28	699	20.9	656	12	US-10-425-114-53846
29	699	20.9	673	12	US-10-425-114-64380
30	693.5	20.7	652	9	US-09-837-992-1
31	693.5	20.7	652	10	US-09-989-981A-2
32	693	20.7	646	14	US-10-079-087-2
33	683.5	20.4	599	15	US-10-210-130-14
34	682.5	20.4	651	9	US-09-837-992-3
35	682.5	20.4	651	10	US-09-989-981A-6
36	682.5	20.4	651	14	US-10-090-455-6
37	677	20.2	559	15	US-10-369-493-5740
38	664	19.8	608	15	US-10-369-493-5748
39	658.5	19.6	1095	15	US-10-369-493-2025
40	657.5	19.6	672	10	US-09-989-981A-4
41	644	19.2	623	12	US-10-424-593-154459
42	640.5	19.1	673	10	US-09-989-981A-8
43	639.5	19.1	673	14	US-10-090-455-7
44	639	19.1	658	15	US-10-369-493-5347
45	636.5	19.0	639	15	US-10-369-493-6184

ALIGNMENTS

RESULT 1

US-09-961-086-1  
Sequence 1, Application US/09961086  
Publication No. US20030036645A1  
GENERAL INFORMATION:  
APPLICANT: UNIVERSITY OF MARYLAND, BALTIMORE  
APPLICANT: ROSS, Douglas D.  
APPLICANT: DOYLE, L. Austin  
APPLICANT: ABRUZZO, Lynne  
TITLE OF INVENTION: BREAST CANCER RESISTANCE PROTEIN (BCRP) AND THE DNA  
TITLE OF INVENTION: WHICH ENCODES IT  
FILE REFERENCE: EP19376-019  
CURRENT APPLICATION NUMBER: US/09/961,086  
PRIOR FILING DATE: 2001-09-21  
PRIOR APPLICATION NUMBER: US 60/073,763  
PRIOR FILING DATE: 1998-02-05  
PRIOR APPLICATION NUMBER: PCT/US99/02577  
PRIOR FILING DATE: 1999-02-05  
NUMBER OF SEQ ID NOS: 7  
SOFTWARE: Patent in ver. 2.1  
SEQ ID NO 1  
LENGTH: 655  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-961-086-1

Query Match 100.0%; Score 3352; DB 10; Length 655;  
Best Local Similarity 100.0%; Pred. No. 7e-300;  
Matches 655; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MSSNVVEFIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICVRVCLKSGFLPCRKPE 60  
Db 1 MSSNVVEFIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICVRVCLKSGFLPCRKPE 60  
QY 61 KEILSNINGIMKPGINAILGPTGGKSSLLDLVLAARXDPGSLGSDVLINGAPRANPKCN 120  
Db 61 KEILSNINGIMKPGINAILGPTGGKSSLLDLVLAARXDPGSLGSDVLINGAPRANPKCN 120

QY 121 SGVVQDDVVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIOELGDKVADSKVGT 180  
Db 121 SGVVQDDVVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIOELGDKVADSKVGT 180  
QY 181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSDSTANAVLLLLKRMKSKQGRITIF 240  
Db 181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSDSTANAVLLLLKRMKSKQGRITIF 240  
QY 241 SIHQPRYSIFKLFDSITLLASGRLMFGHPAQBALGYFESAGYHCEAYNNPADFFLDIING 300  
Db 241 SIHQPRYSIFKLFDSITLLASGRLMFGHPAQBALGYFESAGYHCEAYNNPADFFLDIING 300  
QY 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIEYVNSSFYKETAELHQLSGEKKKK 360  
Db 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIEYVNSSFYKETAELHQLSGEKKKK 360  
QY 361 ITVFEKISYTTGCHQLRWVSKRSFKNLGNPQASIAQIIVTVLGLVIGAIYFGLKND 420  
Db 361 ITVFEKISYTTGCHQLRWVSKRSFKNLGNPQASIAQIIVTVLGLVIGAIYFGLKND 420  
QY 421 TGIONRAGVLFLLTNQCFSSYSAVELFVVEKKLFIHEYISGYRVSSYFGLKLLSDLLP 480  
Db 421 TGIONRAGVLFLLTNQCFSSYSAVELFVVEKKLFIHEYISGYRVSSYFGLKLLSDLLP 480  
QY 481 MTMLPSIIFTCIYVFMGLKPKADAFVMMFTLMVAYSSASSMALAIAAGQSVSVATLL 540  
Db 481 MTMLPSIIFTCIYVFMGLKPKADAFVMMFTLMVAYSSASSMALAIAAGQSVSVATLL 540  
QY 541 MTICFVFMIFSGLLVNLTTIASWLSWQYFSIPRYGFTALQHNFLGQFCPGLNATGN 600  
Db 541 MTICFVFMIFSGLLVNLTTIASWLSWQYFSIPRYGFTALQHNFLGQFCPGLNATGN 600  
QY 601 NPCNATCTGEEYLVKQGDIDSPWGLWKNHVALACMIVIFLTIAVLLKLLFLKKYS 655  
Db 601 NPCNATCTGEEYLVKQGDIDSPWGLWKNHVALACMIVIFLTIAVLLKLLFLKKYS 655

## RESULT 2

US-10-405-806-13  
; Sequence 13, Application US/10405806  
; Publication No. US20030232362A1  
; GENERAL INFORMATION:  
; APPLICANT: KOMATANI, HIDEYA  
; APPLICANT: KARA, YOSHIKAZU  
; APPLICANT: KOTANI, HIDEHITO  
; APPLICANT: NAKAGAWA, RINAKO  
; TITLE OF INVENTION: DRUG RESISTANT GENE AND USE THEREOF  
; FILE REFERENCE: 234985USOCONT  
; CURRENT APPLICATION NUMBER: US/10/405,806  
; CURRENT FILING DATE: 2003-04-03  
; PRIOR APPLICATION NUMBER: PCT/JP01/08112  
; PRIOR FILING DATE: 2001-09-18  
; PRIOR APPLICATION NUMBER: JP2000-303441  
; PRIOR FILING DATE: 2000-10-03  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: Patent in version 3.2  
; SEQ ID NO 13  
; TYPE: PRT  
; LENGTH: 655  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: ABC02 482Mutant sequence  
US-10-405-806-13

Query Match 100.0%; Score 3352; DB 15; Length 655;  
Best Local Similarity 100.0%; Pred. No. 7e-300;  
Matches 655; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MSSSNVEVFIPVSGQNTNGPPATASNDLKAFTEGAVLSFHNIYRVKLSGFLPCRKPVE 60  
Db 1 MSSSNVEVFIPVSGQNTNGPPATASNDLKAFTEGAVLSFHNIYRVKLSGFLPCRKPVE 60  
QY 61 KEILSNINGIMKPLGNAIIGPTGGKSSLLDVLAAKDPGSLGSDVLINGAPRANFKCN 120

Db 61 KEILSNINGIMKPLGNAIIGPTGGKSSLLDVLAAKDPGSLGSDVLINGAPRANFKCN 120  
QY 121 SGVVQDDVVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIOELGDKVADSKVGT 180  
Db 121 SGVVQDDVVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIOELGDKVADSKVGT 180  
QY 181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSDSTANAVLLLLKRMKSKQGRITIF 240  
Db 181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSDSTANAVLLLLKRMKSKQGRITIF 240  
QY 241 SIHQPRYSIFKLFDSITLLASGRLMFGHPAQBALGYFESAGYHCEAYNNPADFFLDIING 300  
Db 241 SIHQPRYSIFKLFDSITLLASGRLMFGHPAQBALGYFESAGYHCEAYNNPADFFLDIING 300  
QY 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIEYVNSSFYKETAELHQLSGEKKKK 360  
Db 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIEYVNSSFYKETAELHQLSGEKKKK 360  
QY 361 ITVFEKISYTTGCHQLRWVSKRSFKNLGNPQASIAQIIVTVLGLVIGAIYFGLKND 420  
Db 361 ITVFEKISYTTGCHQLRWVSKRSFKNLGNPQASIAQIIVTVLGLVIGAIYFGLKND 420  
QY 421 TGIONRAGVLFLLTNQCFSSYSAVELFVVEKKLFIHEYISGYRVSSYFGLKLLSDLLP 480  
Db 421 TGIONRAGVLFLLTNQCFSSYSAVELFVVEKKLFIHEYISGYRVSSYFGLKLLSDLLP 480  
QY 481 MTMLPSIIFTCIYVFMGLKPKADAFVMMFTLMVAYSSASSMALAIAAGQSVSVATLL 540  
Db 481 MTMLPSIIFTCIYVFMGLKPKADAFVMMFTLMVAYSSASSMALAIAAGQSVSVATLL 540  
QY 541 MTICFVFMIFSGLLVNLTTIASWLSWQYFSIPRYGFTALQHNFLGQFCPGLNATGN 600  
Db 541 MTICFVFMIFSGLLVNLTTIASWLSWQYFSIPRYGFTALQHNFLGQFCPGLNATGN 600  
QY 601 NPCNATCTGEEYLVKQGDIDSPWGLWKNHVALACMIVIFLTIAVLLKLLFLKKYS 655  
Db 601 NPCNATCTGEEYLVKQGDIDSPWGLWKNHVALACMIVIFLTIAVLLKLLFLKKYS 655

## RESULT 3

US-09-981-353-35  
; Sequence 35, Application US/09981353  
; Patent No. US20020160382A1  
; GENERAL INFORMATION:  
; APPLICANT: Lasek, Amy W.  
; APPLICANT: Jones, David A.  
; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER  
; FILE REFERENCE: PA-0038 US  
; CURRENT APPLICATION NUMBER: US/09/981,353  
; CURRENT FILING DATE: 2001-10-11  
; NUMBER OF SEQ ID NOS: 194  
; SOFTWARE: PERL Program  
; SEQ ID NO 35  
; LENGTH: 655  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. US20020160382A1 5517972CD1  
US-09-981-353-35

Query Match 99.8%; Score 3346; DB 9; Length 655;  
Best Local Similarity 99.8%; Pred. No. 2.5e-299;  
Matches 654; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSSSNVEVFIPVSGQNTNGPPATASNDLKAFTEGAVLSFHNIYRVKLSGFLPCRKPVE 60  
Db 1 MSSSNVEVFIPVSGQNTNGPPATASNDLKAFTEGAVLSFHNIYRVKLSGFLPCRKPVE 60  
QY 61 KEILSNINGIMKPLGNAIIGPTGGKSSLLDVLAAKDPGSLGSDVLINGAPRANFKCN 120  
Db 61 KEILSNINGIMKPLGNAIIGPTGGKSSLLDVLAAKDPGSLGSDVLINGAPRANFKCN 120

121 SGVYQDDVVMGTLTVRENLQFSAALRLATTMTNHEKNERINRVIOELGLDKVADSKVGT 180  
 121 SGVYQDDVVMGTLTVRENLQFSAALRLATTMTNHEKNERINRVIOELGLDKVADSKVGT 180  
 181 QFIRGSGGERKRTSIGMELITDPSILFDEPTTGLDSDSTANAVALLLKRMSKQGRITIF 240  
 181 QFIRGSGGERKRTSIGMELITDPSILFDEPTTGLDSDSTANAVALLLKRMSKQGRITIF 240  
 241 SIHQPRYSIFKLDLSLTLASGRLMFHPGPAQALGYFESAGVHCEAYNNPADFFLDIING 300  
 241 SIHQPRYSIFKLDLSLTLASGRLMFHPGPAQALGYFESAGVHCEAYNNPADFFLDIING 300  
 301 DSTAVANREEDPKATEIIEPSKQDKPLTEKLAIEIVNSSFYKETAELHOLSGGKXXX 360  
 301 DSTAVANREEDPKATEIIEPSKQDKPLTEKLAIEIVNSSFYKETAELHOLSGGKXXX 360  
 361 ITVFKEISYTTSFCHOLRWVSKRSFNKLLGNPOASIAQIIVTVLGLVIGAIYFGLKND 420  
 361 ITVFKEISYTTSFCHOLRWVSKRSFNKLLGNPOASIAQIIVTVLGLVIGAIYFGLKND 420  
 421 TGIONRAGVLFLTTNQCFSSVSAVELFVVEKKLFIHEYISGYRVSSYFGLKLLSLLP 480  
 421 TGIONRAGVLFLTTNQCFSSVSAVELFVVEKKLFIHEYISGYRVSSYFGLKLLSLLP 480  
 481 MTMLPSIIFTCIIVFVLMGLKPKADAFVVMFTLMWVAYSASSMALAIAAGQSVSVATLL 540  
 481 MRMLPSIIFTCIIVFVLMGLKPKADAFVVMFTLMWVAYSASSMALAIAAGQSVSVATLL 540  
 541 MTICFVFMIFSGLLVNLTTIASWSLQYFSIPRYGFTALQHNFLGQFCPLNATGN 600  
 541 MTICFVFMIFSGLLVNLTTIASWSLQYFSIPRYGFTALQHNFLGQFCPLNATGN 600  
 601 NPCNYATCTGEEYLVKQIDLSPWGLWKNHVALACMIVFLTIAYLKLLFLKXXS 655  
 601 NPCNYATCTGEEYLVKQIDLSPWGLWKNHVALACMIVFLTIAYLKLLFLKXXS 655

RESULT 4  
 US-10-120-687-61  
 ; Sequence 61, Application US/10120687  
 ; Publication No. US20030082155A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Massachusetts General Hospital  
 ; TITLE OF INVENTION: Stem Cells of the Islets of Langerhans and Their Use in Treating  
 ; FILE REFERENCE: 3284/1235B  
 ; CURRENT APPLICATION NUMBER: US/10/120,687  
 ; CURRENT FILING DATE: 2002-04-11  
 ; PRIOR APPLICATION NUMBER: US60/169082  
 ; PRIOR FILING DATE: 1999-12-06  
 ; PRIOR APPLICATION NUMBER: US 09/963,875  
 ; PRIOR FILING DATE: 2001-09-25  
 ; PRIOR APPLICATION NUMBER: US 60/215109  
 ; PRIOR FILING DATE: 2000-06-28  
 ; PRIOR APPLICATION NUMBER: US 60/238890  
 ; PRIOR FILING DATE: 2000-10-06  
 ; PRIOR APPLICATION NUMBER: US 09/731261  
 ; PRIOR FILING DATE: 2000-12-06  
 ; NUMBER OF SEQ ID NOS: 61  
 ; SOFTWARE: Patent in version 3.1  
 ; SEQ ID NO 61  
 ; LENGTH: 655  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-10-120-687-61

Query Match 99.8%; Score 3346; DB 14; Length 655;  
 Best Local Similarity 99.8%; Pred. No. 2.5e-299;  
 Matches 654; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 1 MSSNVEVFPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPVE 60

1 MSSNVEVFPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPVE 60  
 61 KEILSNINGIMKPLGNAIIGPTGGKSSLLDVLAAKDPGLSGDVLINGAPRANFKCN 120  
 61 KEILSNINGIMKPLGNAIIGPTGGKSSLLDVLAAKDPGLSGDVLINGAPRANFKCN 120  
 121 SGVYQDDVVMGTLTVRENLQFSAALRLATTMTNHEKNERINRVIOELGLDKVADSKVGT 180  
 121 SGVYQDDVVMGTLTVRENLQFSAALRLATTMTNHEKNERINRVIOELGLDKVADSKVGT 180  
 181 QFIRGSGGERKRTSIGMELITDPSILFDEPTTGLDSDSTANAVALLLKRMSKQGRITIF 240  
 181 QFIRGSGGERKRTSIGMELITDPSILFDEPTTGLDSDSTANAVALLLKRMSKQGRITIF 240  
 241 SIHQPRYSIFKLDLSLTLASGRLMFHPGPAQALGYFESAGVHCEAYNNPADFFLDIING 300  
 241 SIHQPRYSIFKLDLSLTLASGRLMFHPGPAQALGYFESAGVHCEAYNNPADFFLDIING 300  
 301 DSTAVANREEDPKATEIIEPSKQDKPLTEKLAIEIVNSSFYKETAELHOLSGGKXXX 360  
 301 DSTAVANREEDPKATEIIEPSKQDKPLTEKLAIEIVNSSFYKETAELHOLSGGKXXX 360  
 361 ITVFKEISYTTSFCHOLRWVSKRSFNKLLGNPOASIAQIIVTVLGLVIGAIYFGLKND 420  
 361 ITVFKEISYTTSFCHOLRWVSKRSFNKLLGNPOASIAQIIVTVLGLVIGAIYFGLKND 420  
 421 TGIONRAGVLFLTTNQCFSSVSAVELFVVEKKLFIHEYISGYRVSSYFGLKLLSLLP 480  
 421 TGIONRAGVLFLTTNQCFSSVSAVELFVVEKKLFIHEYISGYRVSSYFGLKLLSLLP 480  
 481 MTMLPSIIFTCIIVFVLMGLKPKADAFVVMFTLMWVAYSASSMALAIAAGQSVSVATLL 540  
 481 MRMLPSIIFTCIIVFVLMGLKPKADAFVVMFTLMWVAYSASSMALAIAAGQSVSVATLL 540  
 541 MTICFVFMIFSGLLVNLTTIASWSLQYFSIPRYGFTALQHNFLGQFCPLNATGN 600  
 541 MTICFVFMIFSGLLVNLTTIASWSLQYFSIPRYGFTALQHNFLGQFCPLNATGN 600  
 601 NPCNYATCTGEEYLVKQIDLSPWGLWKNHVALACMIVFLTIAYLKLLFLKXXS 655  
 601 NPCNYATCTGEEYLVKQIDLSPWGLWKNHVALACMIVFLTIAYLKLLFLKXXS 655

Query Match 99.8%; Score 3346; DB 15; Length 655;  
 Best Local Similarity 99.8%; Pred. No. 2.5e-299;  
 Matches 654; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 1 MSSNVEVFPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPVE 60

```
Db      1  MSSSNVEFIPVQNTNGFPATASNDLKAFTEGAVLSFHNI CYRVKLSGFLPCRKPE 60
Qy      61  KEILSNINGIMKPGNALIGPTGGKSSLDVLAARKDPSGLSGDVLINGAPRANFKCN 120
Db      61  KEILSNINGIMKPGNALIGPTGGKSSLDVLAARKDPSGLSGDVLINGAPRANFKCN 120
Qy     121  SGVYVQDDVVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIQELGDKVADSKVGT 180
Db     121  SGVYVQDDVVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIQELGDKVADSKVGT 180
Qy     181  QFIRGVSGGERKRTSIGMELITDPSILFDEPTTGLDSDSTANAVALLKRWMSKQRTIIF 240
Db     181  QFIRGVSGGERKRTSIGMELITDPSILFDEPTTGLDSDSTANAVALLKRWMSKQRTIIF 240
Qy     241  SIHQPRYSIFKLFDSLTLASGRLMFHGPAQALGYFESAGYHCEAYNNPADFFLDIING 300
Db     241  SIHQPRYSIFKLFDSLTLASGRLMFHGPAQALGYFESAGYHCEAYNNPADFFLDIING 300
Qy     301  DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIEIYVNSSFYKETAELHQLSGEKKKK 360
Db     301  DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIEIYVNSSFYKETAELHQLSGEKKKK 360
Qy     361  ITVFXEISYTTSFCHQLRWVSKRSFKNLGNPQASIAQIIVTVVLGLVIGAIYFGLKXDS 420
Db     361  ITVFXEISYTTSFCHQLRWVSKRSFKNLGNPQASIAQIIVTVVLGLVIGAIYFGLKXDS 420
Qy     421  TGIQNRAGVLPFLTNQCPSSVSAYVELFVVEKKLFIHEYISGYRVSSYFGLKLLSDLLP 480
Db     421  TGIQNRAGVLPFLTNQCPSSVSAYVELFVVEKKLFIHEYISGYRVSSYFGLKLLSDLLP 480
Qy     481  MTMLPSIIFTCTIVYFVLMGLKPKADAFFVVMFTLMVAVSASSMALAIAAGQSVSVATLL 540
Db     481  MTMLPSIIFTCTIVYFVLMGLKPKADAFFVVMFTLMVAVSASSMALAIAAGQSVSVATLL 540
Qy     541  MTICFVFMVIFSGLLVNLTTIASWLSWQYFSIPRYGFTALQHNBFQCNFCPLNATGN 600
Db     541  MTICFVFMVIFSGLLVNLTTIASWLSWQYFSIPRYGFTALQHNBFQCNFCPLNATGN 600
Qy     601  NPCNVATCTGEEYLVKQGDLSFPGWLNKHNVALACMIVIFLTIAVLKLLFLKKYS 655
Db     601  NPCNVATCTGEEYLVKQGDLSFPGWLNKHNVALACMIVIFLTIAVLKLLFLKKYS 655
```

## RESULT 6

```
US-09-866-866A-27
; Sequence 27, Application US/09866866A
; Patent No. US20020102244A1
; GENERAL INFORMATION:
; APPLICANT: Sorrentino, Brian
; APPLICANT: Schuetz, John
; TITLE OF INVENTION: A Method of Identifying and/or Isolating Stem Cells
; FILE REFERENCE: 1340-1-021CIP2
; CURRENT APPLICATION NUMBER: US/09/866,866A
; CURRENT FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: 09/584,586
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: PCT/US99/11825
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: 60/086,988
; PRIOR FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 27
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-866-866A-27
```

Query Match 99.6%; Score 3338; DB 9; Length 655;  
Best Local Similarity 99.7%; Pred. No. 1.4e-298; Indels 0; Gaps 0;  
Matches 653; Conservative 0; Mismatches 2;

```
Qy     1  MSSSNVEFIPVQNTNGFPATASNDLKAFTEGAVLSFHNI CYRVKLSGFLPCRKPE 60
Db     1  MSSSNVEFIPVQNTNGFPATASNDLKAFTEGAVLSFHNI CYRVKLSGFLPCRKPE 60
Qy     61  KEILSNINGIMKPGNALIGPTGGKSSLDVLAARKDPSGLSGDVLINGAPRANFKCN 120
Db     61  KEILSNINGIMKPGNALIGPTGGKSSLDVLAARKDPSGLSGDVLINGAPRANFKCN 120
Qy     121  SGVYVQDDVVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIQELGDKVADSKVGT 180
Db     121  SGVYVQDDVVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIQELGDKVADSKVGT 180
Qy     181  QFIRGVSGGERKRTSIGMELITDPSILFDEPTTGLDSDSTANAVALLKRWMSKQRTIIF 240
Db     181  QFIRGVSGGERKRTSIGMELITDPSILFDEPTTGLDSDSTANAVALLKRWMSKQRTIIF 240
Qy     241  SIHQPRYSIFKLFDSLTLASGRLMFHGPAQALGYFESAGYHCEAYNNPADFFLDIING 300
Db     241  SIHQPRYSIFKLFDSLTLASGRLMFHGPAQALGYFESAGYHCEAYNNPADFFLDIING 300
Qy     301  DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIEIYVNSSFYKETAELHQLSGEKKKK 360
Db     301  DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIEIYVNSSFYKETAELHQLSGEKKKK 360
Qy     361  ITVFXEISYTTSFCHQLRWVSKRSFKNLGNPQASIAQIIVTVVLGLVIGAIYFGLKXDS 420
Db     361  ITVFXEISYTTSFCHQLRWVSKRSFKNLGNPQASIAQIIVTVVLGLVIGAIYFGLKXDS 420
Qy     421  TGIQNRAGVLPFLTNQCPSSVSAYVELFVVEKKLFIHEYISGYRVSSYFGLKLLSDLLP 480
Db     421  TGIQNRAGVLPFLTNQCPSSVSAYVELFVVEKKLFIHEYISGYRVSSYFGLKLLSDLLP 480
Qy     481  MTMLPSIIFTCTIVYFVLMGLKPKADAFFVVMFTLMVAVSASSMALAIAAGQSVSVATLL 540
Db     481  MTMLPSIIFTCTIVYFVLMGLKPKADAFFVVMFTLMVAVSASSMALAIAAGQSVSVATLL 540
Qy     541  MTICFVFMVIFSGLLVNLTTIASWLSWQYFSIPRYGFTALQHNBFQCNFCPLNATGN 600
Db     541  MTICFVFMVIFSGLLVNLTTIASWLSWQYFSIPRYGFTALQHNBFQCNFCPLNATGN 600
Qy     601  NPCNVATCTGEEYLVKQGDLSFPGWLNKHNVALACMIVIFLTIAVLKLLFLKKYS 655
Db     601  NPCNVATCTGEEYLVKQGDLSFPGWLNKHNVALACMIVIFLTIAVLKLLFLKKYS 655
```

## RESULT 7

```
US-09-866-866A-10
; Sequence 10, Application US/09866866A
; Patent No. US20020102244A1
; GENERAL INFORMATION:
; APPLICANT: Sorrentino, Brian
; APPLICANT: Schuetz, John
; TITLE OF INVENTION: A Method of Identifying and/or Isolating Stem Cells
; FILE REFERENCE: 1340-1-021CIP2
; CURRENT APPLICATION NUMBER: US/09/866,866A
; CURRENT FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: 09/584,586
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: PCT/US99/11825
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: 60/086,988
; PRIOR FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-866-866A-10
```

Query Match 99.4%; Score 3331; DB 9; Length 655;  
Best Local Similarity 99.4%; Pred. No. 6.1e-298; Indels 0; Gaps 0;  
Matches 651; Conservative 1; Mismatches 3;

QY 1 MSSNVEFIPVQSGNTGFPATASNDLKAFTEGAVLSFHNIYRVKLSGFLPCRKPE 60  
Db 1 MSSNVEFIPVQSGNTGFPATVNDLKAFTEGAVLSFHNIYRVKLSGFLPCRKPE 60  
QY 61 KEILSNINGIMKPGNAILGPTGGKSSLLDLAARKOPSGLSGVDVINGAPRANFKCN 120  
Db 61 KEILSNINGIMKPGNAILGPTGGKSSLLDLAARKOPSGLSGVDVINGAPRANFKCN 120  
QY 121 SGVVODDVMGTLTVRENLQPSAALRLATTNHEKNERINRVIOELGLDKVADSKVGT 180  
Db 121 SGVVODDVMGTLTVRENLQPSAALRLATTNHEKNERINRVIOELGLDKVADSKVGT 180  
QY 181 QTRGVSGGERKRTSIGMELITDPSILFDEPTTGLDSTANAVLLLLKRMKSQKRTIIF 240  
Db 181 QTRGVSGGERKRTSIGMELITDPSILFDEPTTGLDSTANAVLLLLKRMKSQKRTIIF 240  
QY 241 SHQPRYSIFKLPDSLTLLASGRMLPHGPAQAEALGYFESAGYHCEAYNNPADFFLDIING 300  
Db 241 SHQPRYSIFKLPDSLTLLASGRMLPHGPAQAEALGYFESAGYHCEAYNNPADFFLDIING 300  
QY 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIEIYVNSFYKETAELHQLSGGKKKK 360  
Db 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIEIYVNSFYKETAELHQLSGGKKKK 360  
QY 361 ITVFKESIYTTSPCHQLRWYSKSFKNLLGNPOASTAQIIVTVVLGIVGAIYFGLKND 420  
Db 361 ITVFKESIYTTSPCHQLRWYSKSFKNLLGNPOASTAQIIVTVVLGIVGAIYFGLKND 420  
QY 421 TGIONRAGVLPFLTTNOCFSSVSAVELFVVEKKLFIHEYISGYRVSSYFGLKLLSLLP 480  
Db 421 TGIONRAGVLPFLTTNOCFSSVSAVELFVVEKKLFIHEYISGYRVSSYFGLKLLSLLP 480  
QY 481 MTMLPSIIFTCIVYFVLMGLPKADAFVVMFTLMVAYASASSMALAIAAGQSVSVATLL 540  
Db 481 MRMLPSIIFTCIVYFVLMGLPKADAFVVMFTLMVAYASASSMALAIAAGQSVSVATLL 540  
QY 541 MTICFVPMIIFSGLLVNLTTIASWLSWLOYSIPRYGFTALQHNFLQNFCEGLNATGN 600  
Db 541 MTICFVPMIIFSGLLVNLTTIASWLSWLOYSIPRYGFTALQHNFLQNFCEGLNATGN 600  
QY 601 NPCNYATCTGEBYLVKQIDLSPLWGLMKNHVALACMIVIFLTITAYLKLLFLKKYS 655  
Db 601 NPCNYATCTGEBYLVKQIDLSPLWGLMKNHVALACMIVIFLTITAYLKLLFLKKYS 655

RESULT 8  
US-10-090-455-5  
; Sequence 5, Application US/10090455  
; Publication No. US20030027259A1  
; GENERAL INFORMATION:  
; APPLICANT: Chen, Hongyun  
; APPLICANT: Le Bihan, Stephane  
; TITLE OF INVENTION: NOVEL ABCG4 TRANSPORTER AND USES THEREOF  
; FILE REFERENCE: 100103.406  
; CURRENT APPLICATION NUMBER: US/10/090.455  
; CURRENT FILING DATE: 2002-03-01  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 655  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-090-455-5  
Query Match 99.4%; Score 3331; DB 14; Length 655;  
Best Local Similarity 99.4%; Pred. No. 6.1e-298;  
Matches 651; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSSNVEFIPVQSGNTGFPATASNDLKAFTEGAVLSFHNIYRVKLSGFLPCRKPE 60  
Db 1 MSSNVEFIPVQSGNTGFPATVNDLKAFTEGAVLSFHNIYRVKLSGFLPCRKPE 60

QY 61 KEILSNINGIMKPGNAILGPTGGKSSLLDLAARKOPSGLSGVDVINGAPRANFKCN 120  
Db 61 KEILSNINGIMKPGNAILGPTGGKSSLLDLAARKOPSGLSGVDVINGAPRANFKCN 120  
QY 121 SGVVODDVMGTLTVRENLQPSAALRLATTNHEKNERINRVIOELGLDKVADSKVGT 180  
Db 121 SGVVODDVMGTLTVRENLQPSAALRLATTNHEKNERINRVIOELGLDKVADSKVGT 180  
QY 181 QTRGVSGGERKRTSIGMELITDPSILFDEPTTGLDSTANAVLLLLKRMKSQKRTIIF 240  
Db 181 QTRGVSGGERKRTSIGMELITDPSILFDEPTTGLDSTANAVLLLLKRMKSQKRTIIF 240  
QY 241 SHQPRYSIFKLPDSLTLLASGRMLPHGPAQAEALGYFESAGYHCEAYNNPADFFLDIING 300  
Db 241 SHQPRYSIFKLPDSLTLLASGRMLPHGPAQAEALGYFESAGYHCEAYNNPADFFLDIING 300  
QY 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIEIYVNSFYKETAELHQLSGGKKKK 360  
Db 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIEIYVNSFYKETAELHQLSGGKKKK 360  
QY 361 ITVFKESIYTTSPCHQLRWYSKSFKNLLGNPOASTAQIIVTVVLGIVGAIYFGLKND 420  
Db 361 ITVFKESIYTTSPCHQLRWYSKSFKNLLGNPOASTAQIIVTVVLGIVGAIYFGLKND 420  
QY 421 TGIONRAGVLPFLTTNOCFSSVSAVELFVVEKKLFIHEYISGYRVSSYFGLKLLSLLP 480  
Db 421 TGIONRAGVLPFLTTNOCFSSVSAVELFVVEKKLFIHEYISGYRVSSYFGLKLLSLLP 480  
QY 481 MTMLPSIIFTCIVYFVLMGLPKADAFVVMFTLMVAYASASSMALAIAAGQSVSVATLL 540  
Db 481 MRMLPSIIFTCIVYFVLMGLPKADAFVVMFTLMVAYASASSMALAIAAGQSVSVATLL 540  
QY 541 MTICFVPMIIFSGLLVNLTTIASWLSWLOYSIPRYGFTALQHNFLQNFCEGLNATGN 600  
Db 541 MTICFVPMIIFSGLLVNLTTIASWLSWLOYSIPRYGFTALQHNFLQNFCEGLNATGN 600  
QY 601 NPCNYATCTGEBYLVKQIDLSPLWGLMKNHVALACMIVIFLTITAYLKLLFLKKYS 655  
Db 601 NPCNYATCTGEBYLVKQIDLSPLWGLMKNHVALACMIVIFLTITAYLKLLFLKKYS 655

RESULT 9  
US-09-745-763-197  
; Sequence 197, Application US/09745763  
; Patent No. US20020065394A1  
; GENERAL INFORMATION:  
; APPLICANT: Jacobs, Kenneth  
; McCoy, John M.  
; LaVallie, Edward R.  
; Collins-Racie, Lisa A.  
; Svans, Cheryl  
; Merberg, David  
; Treacy, Maurice  
; Spaulding, Vikki  
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES  
; NUMBER OF SEQUENCES: 219  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 Cambridgepark Drive  
; CITY: Cambridge  
; STATE: MA  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/745,763  
; FILING DATE: 18-Jun-2000  
; CLASSIFICATION: <Unknown>



## ATTORNEY/AGENT INFORMATION:

NAME: Sprunger, Suzanne A.  
REGISTRATION NUMBER: 41,323  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 498-8284  
TELEFAX: (617) 876-5851

## INFORMATION FOR SEQ ID NO: 197:

SEQUENCE CHARACTERISTICS:  
LENGTH: 604 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 197:  
US-09-745-763-197

Query Match 91.1%; Score 3053.5; DB 9; Length 604;  
Best Local Similarity 91.9%; Pred. No. 2,1e-272;  
Matches 602; Conservative 0; Mismatches 2; Indels 51; Gaps 1;

QY 1 MSSNVVEFIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICRYVKLSGFLPCRPVE 60  
DB 1 MSSNVVEFIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICRYVKLSGFLPCRPVE 60

QY 61 KEILSNINGIMKPGNALIGPTGGKSSLLDLVLAARKDPSGLSGDVLINGAPRANPKCN 120  
DB 61 KEILSNINGIMKPGNALIGPTGGKSSLLDLVLAARKDPSGLSGDVLINGAPRANPKCN 120

QY 121 SGYVQDDVVMGTLTVRENLFQSAALRLATTWNHEKNERINRVIOELGLDKVADSKVGT 180  
DB 121 SGYVQDDVVMGTLTVRENLFQSAALRLATTWNHEKNERINRVIOELGLDKVADSKVGT 180

QY 181 QTRGVSGGERKRTSIGMELITDPSILFLDEPTTGLSDSTANAVALLKRMKSKQRTIIF 240  
DB 181 QTRGVSGGERKRTSIGMELITDPSILFLDEPTTGLSDSTANAVALLKRMKSKQRTIIF 240

QY 241 SIHOPRYSIFKLPDSLTLLASGLRMFHPGAQALGYFESAGYHCEAYNNPADFFLDIING 300  
DB 241 SIHOPRYSIFKLPDSLTLLASGLRMFHPGAQALGYFESAGYHCEAYNNPADFFLDIING 300

QY 301 DSTAVALNEEDPKATEIIEPSKQDKPLIEKLAETVYNSSFYKETAELHOLSGGEKKK 360  
DB 250 DSTAVALNEEDPKATEIIEPSKQDKPLIEKLAETVYNSSFYKETAELHOLSGGEKKK 309

QY 361 ITVFKEISYTTSFCHQLRWKRSFKNLLGNPQASIAQIIVTVLGLVIGAIYFGLKND 420  
DB 310 ITVFKEISYTTSFCHQLRWKRSFKNLLGNPQASIAQIIVTVLGLVIGAIYFGLKND 369

QY 421 TGIONRAGVLPFLTTNQCFSSVSAYELFVVEKKLFIHEYISGYRVSSYFGLKLSDLLP 480  
DB 370 TGIONRAGVLPFLTTNQCFSSVSAYELFVVEKKLFIHEYISGYRVSSYFGLKLSDLLP 429

QY 481 MTMLPSIIFTCTIVYFVLMGLKPKADAFFVMMFTLMMVAYSASSMALAIAGQSVSVATLL 540  
DB 430 MTMLPSIIFTCTIVYFVLMGLKPKADAFFVMMFTLMMVAYSASSMALAIAGQSVSVATLL 489

QY 541 MTICFVFMFI FSGLLVNLTTIASWSLQYFIPRYGFTALOHNEFLQNCPCPLNATGN 600  
DB 490 MTICFVFMFI FSGLLVNLTTIASWSLQYFIPRYGFTALOHNEFLQNCPCPLNATGN 549

QY 601 NPCNVATCTGEEYLVKQIDLSFWGLWKNHVALACMIVIFLTIAVLLKLLFLKKYS 655  
DB 550 NPCNVATCTGEEYLVKQIDLSFWGLWKNHVALACMIVIFLTIAVLLKLLFLKKYS 604

## RESULT 10

US-09-866-866A-14  
Sequence 14, Application US/09866866A  
Patent No. US20020102244A1  
GENERAL INFORMATION:  
APPLICANT: Sorrentino, Brian  
APPLICANT: Schuetz, John  
TITLE OF INVENTION: A Method of Identifying and/or Isolating Stem Cells

## FILE REFERENCE: 1340-1-021CIP2

CURRENT APPLICATION NUMBER: US/09/866,866A  
CURRENT FILING DATE: 2001-08-30  
PRIOR APPLICATION NUMBER: 09/584,586  
PRIOR FILING DATE: 2000-05-31  
PRIOR APPLICATION NUMBER: PCI/US99/11825  
PRIOR FILING DATE: 1999-05-27  
PRIOR APPLICATION NUMBER: 60/086,988  
PRIOR FILING DATE: 1998-05-28  
NUMBER OF SEQ ID NOS: 27  
SOFTWARE: Patentin version 3.0  
SEQ ID NO 14  
LENGTH: 657  
TYPE: PRT  
ORGANISM: Mus musculus  
US-09-866-866A-14

Query Match 82.2%; Score 2757; DB 9; Length 657;  
Best Local Similarity 81.5%; Pred. No. 5,5e-245;  
Matches 536; Conservative 51; Mismatches 67; Indels 4; Gaps 3;

QY 1 MSSNVVEFIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICRYVKLSGFLPCRPVE 60  
DB 1 MSSNDHVLVPMQSRNNGLPRMNSRAVRTLAEGDVLSPHHTYRVKVSGLF-VRKTYE 59

QY 61 KEILSNINGIMKPGNALIGPTGGKSSLLDLVLAARKDPSGLSGDVLINGAPRANPKCN 120  
DB 60 KEILSNINGIMKPGNALIGPTGGKSSLLDLVLAARKDPSGLSGDVLINGAPRANPKCN 119

QY 121 SGYVQDDVVMGTLTVRENLFQSAALRLATTWNHEKNERINRVIOELGLDKVADSKVGT 180  
DB 120 SGYVQDDVVMGTLTVRENLFQSAALRLATTWNHEKNERINRVIOELGLDKVADSKVGT 179

QY 181 QTRGVSGGERKRTSIGMELITDPSILFLDEPTTGLSDSTANAVALLKRMKSKQRTIIF 240  
DB 180 QTRGVSGGERKRTSIGMELITDPSILFLDEPTTGLSDSTANAVALLKRMKSKQRTIIF 239

QY 241 SIHOPRYSIFKLPDSLTLLASGLRMFHPGAQALGYFESAGYHCEAYNNPADFFLDIING 300  
DB 240 SIHOPRYSIFKLPDSLTLLASGLRVHPGAQALGYFESAGYHCEAYNNPADFFLDIING 299

QY 301 DSTAVALNEEDPKATEIIEPSKQDKPLIEKLAETVYNSSFYKETAELHOLSGGEKKK 359  
DB 300 DSAVVALNEEDQNEANKTEEPSKGEKPVLENSEFIYNSAIYGETKAELOLPQGEKKK 359

QY 360 KITVFEISYTTSFCHQLRWKRSFKNLLGNPQASIAQIIVTVLGLVIGAIYFGLKND 419  
DB 360 GTSAFKEPVVTSFCHQLRWIARRSFKNLLGNPQASIAQIIVTVLGLVIGAIYFGLKND 419

QY 420 TGIONRAGVLPFLTTNQCFSSVSAYELFVVEKKLFIHEYISGYRVSSYFGLKLSDLL 479  
DB 420 AAGMQNRAGVLPFLTTNQCFSSVSAYELFVVEKKLFIHEYISGYRVSSYFGLKNSDLL 479

QY 480 PMTMLPSIIFTCTIVYFVLMGLKPKADAFFVMMFTLMMVAYSASSMALAIAGQSVSVATLL 539  
DB 480 PMTFPSVIFTCTIVYFVLMGLKKTVDFAFFIMFTLINVAYTASSMALAIATGQSVSVATLL 539

QY 540 LMTICFVFMFI FSGLLVNLTTIASWSLQYFIPRYGFTALOHNEFLQNCPCPLNATGN 599  
DB 540 LMTIAFVFMFI FSGLLVNLRTIGFWLSWQYFIPRYGFTALQYNEFLQGEPCPGFNVD 599

QY 600 NNPC--NYATCTGEEYLVKQIDLSFWGLWKNHVALACMIVIFLTIAVLLKLLFLKKYS 655  
DB 600 NSTCVNSYAICTGNEVLINQGIELSPWGLWKNHVALACMIIIFLTIAVLLKLLFLKKYS 657

## RESULT 11

US-10-369-493-1520  
Sequence 1520, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.



Db 605 EISCTSSN----TTCPSGKGVILETINFSAADPLDYVGLAILIVSRVLAYLAL 655

## RESULT 13

US-10-424-599-242078  
; Sequence 242078, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 242078  
; LENGTH: 1084  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_60623C.1.pap  
US-10-424-599-242078

Query Match 24.0%; Score 805; DB 12; Length 1084;  
Best Local Similarity 29.4%; Pred. No. 1.5e-64;  
Matches 217; Conservative 120; Mismatches 250; Indels 150; Gaps 17;

QY 35 AVLSFHNICVRVKLSGFLPCRKPVEKEILSNINIMKPG-LNAILGPTGGKSLLDVL 93  
Db 373 ASLAFENVSHLK-----GRQILSDISGAVHFGELMAMGASGAGKTTFLDLL 420  
QY 94 AARKDPSGLS-GDVLINGAP-RPANKKNSGVVDDVMGTLTVRENLOFSAALRLATT 151  
Db 421 -ARKNRGTGLGDFWNGEKITDDDFRSAGIFVDDQDMLPTLVHETIMDSALLRLPKD 479  
QY 152 MTNHEKNERINRVIQELGLDKVADSKVGTQ--FIRVSGGGERKRTSIGMELITDPSILFL 209  
Db 480 MSWAVKAQKVEDVEKQIGYHIRHQLIGSESGRGISGGERKRVGIACELVTSFSLFL 539  
QY 210 DEPTTGLDSSA-NAVLLLLKMSKQGTIIFSIHQPRYSIFKLPDSLTLLASGLMPHG 268  
Db 540 YBPTSLGDYAFNVEICLVTLVKYNTVWFTIHQPRSNIVALPDQLVLLAQGTVYSG 599  
QY 269 PAQEAALGPESAGYHCEAYNPNADPFLDI-----INGDSTAVAL 307  
Db 600 PFSKQNYFDSIGYSCPPGFNADYLDVTMHASRPDNPIDPSHTYGEDLDGRATVASS 659  
QY 308 NREEDFKATEIIEPS-----KQDKPLIE----- 330  
Db 660 TMAVKSIPSIISNPSIEGRPKNKRKSIROOQERQLFSRKSTDRAPASLRSDGSIL 719  
QY 331 -----KLA-----EYVNSSFYKETAELH----- 350  
Db 720 DKDNKWKVQARQAQSSINEDGHDLPDPAGNNTNLDVLVSSYEASDVKAELHSDIESAI 779  
QY 351 -----QLSGEKKKKITVKEISYTTSFCHQLRWKSRKSPKLLGNPQASIAQI 399  
Db 780 SDASESNGNPLHAPNGNSGKRGYKPKMLPTQFL-----ILSRTRWNLRYNPQMLTHY 835  
QY 400 IVTVVLGLVIGATYFGLKNDSTGIONRAGVLFELTTNOCFSSVSAVELFVVEKKLFIHEY 459  
Db 836 ATAILLAVPLGLFYGLTDLKGFQNLGFFFLALFGSFLTLTTFAPERLLFVER 895  
QY 460 ISGYRVSSYFGLKSLDLPMTMLPSIIFTCIVTFMLGLKPKADAFPMVMTLMKWAYS 519  
Db 896 AKGYTFLAYLAKVDFDVPLRIPLPAIMGIVYPMTGLVPTAPEFGFKMLLLVFNLA 955  
QY 520 ASSMALIAAGQSVSWATLLMTICFVFMWIFSGLLVNLTTIASLWSLQVFSIPRYGFT 579  
Db 956 AMMICLFLGIVVRNQGVANLLGVLMFLSLFGGLLHETIPKGLLWQLSLIFHFGFE 1015

QY 580 ALQHNE-----FLGQNFCEPLNATGNPCNATCTGEEYLVKQIDLSWGLWKHVALA 634  
Db 1016 GLIVNEVRVLSLVDDKY--GLN-----LEVFGSAILSSFGFDV--LALWRDVAGLG 1062  
QY 635 CMIVIFLTITAYLKLLPL 651  
Db 1063 VVSGAFFVLGFAALHFL 1079

## RESULT 14

US-10-425-114-63125  
; Sequence 63125, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E.  
; APPLICANT: Tabaska, Jack E.  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 63125  
; LENGTH: 1101  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: JC-GMLE01810066F10\_FLI.pap  
US-10-425-114-63125

Query Match 24.0%; Score 805; DB 12; Length 1101;

Best Local Similarity 29.4%; Pred. No. 1.5e-64;  
Matches 217; Conservative 120; Mismatches 250; Indels 150; Gaps 17;

QY 35 AVLSFHNICVRVKLSGFLPCRKPVEKEILSNINIMKPG-LNAILGPTGGKSLLDVL 93  
Db 390 ASLAFENVSHLK-----GRQILSDISGAVHFGELMAMGASGAGKTTFLDLL 437  
QY 94 AARKDPSGLS-GDVLINGAP-RPANKKNSGVVDDVMGTLTVRENLOFSAALRLATT 151  
Db 438 -ARKNRGTGLGDFWNGEKITDDDFRSAGIFVDDQDMLPTLVHETIMDSALLRLPKD 496  
QY 152 MTNHEKNERINRVIQELGLDKVADSKVGTQ--FIRVSGGGERKRTSIGMELITDPSILFL 209  
Db 497 MSWAVKAQKVEDVEKQIGYHIRHQLIGSESGRGISGGERKRVGIACELVTSFSLFL 556  
QY 210 DEPTTGLDSSA-NAVLLLLKMSKQGTIIFSIHQPRYSIFKLPDSLTLLASGLMPHG 268  
Db 557 YBPTSLGDYAFNVEICLVTLVKYNTVWFTIHQPRSNIVALPDQLVLLAQGTVYSG 616  
QY 269 PAQEAALGPESAGYHCEAYNPNADPFLDI-----INGDSTAVAL 307  
Db 617 PFSKQNYFDSIGYSCPPGFNADYLDVTMHASRPDNPIDPSHTYGEDLDGRATVASS 676  
QY 308 NREEDFKATEIIEPS-----KQDKPLIE----- 330  
Db 677 TMAVKSIPSIISNPSIEGRPKNKRKSIROOQERQLFSRKSTDRAPASLRSDGSIL 736  
QY 331 -----KLA-----EYVNSSFYKETAELH----- 350  
Db 737 DKDNKWKVQARQAQSSINEDGHDLPDPAGNNTNLDVLVSSYEASDVKAELHSDIESAI 796  
QY 351 -----QLSGEKKKKITVKEISYTTSFCHQLRWKSRKSPKLLGNPQASIAQI 399  
Db 797 SDASESNGNPLHAPNGNSGKRGYKPKMLPTQFL-----ILSRTRWNLRYNPQMLTHY 852  
QY 400 IVTVVLGLVIGATYFGLKNDSTGIONRAGVLFELTTNOCFSSVSAVELFVVEKKLFIHEY 459

853 ATAILLAVFLGYGLTDDKGFQNLKGFLLALFGFSTLTITFAPERLLFVRER 912  
460 ISGYVRVSSYFLKLLSLLPMTLPSIITFCIVVFMLGLKPKADAFVMMETLMMVAYS 519  
913 AKGYTPLAYYLAQVDFIVLRIUPLPAPMGICIVVPMTGIVTAPDEFKFMILLVFLNLA 972  
520 ASSMALIAAGOSVSVATLLMTICFVFMIFSGLLVNLTTIASLWMLQYFSPRYGFT 579  
973 AAMICFLGIVVNRQVGNLGLVNLVMSLLFGGLLNNHETIPKGLLWLSLSIFHPGFE 1032  
580 ALOHNE-----FLGQNFPCLNATGNNPCYATCTGEEVLYKQGLDLSFWGLKKNHVALA 634  
1033 GLTIVNEVRLSLVDKXY--GLN-----IEVPGSAILSSFGFDV--LALWRDVAGLG 1079  
635 CMIVIELTIAYLKLEL 651  
1080 VWSGAFFVLGFAALHFL 1096  
RESULT 15  
US-10-369-493-3799  
; Sequence 3799, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 3799  
; LENGTH: 676  
; TYPE: PRT  
; ORGANISM: Neurospora crassa  
US-10-369-493-3799  
Query Match 23.7%; Score 795.5; DB 15; Length 676;  
Best Local Similarity 31.2%; Pred. No. 5.4e-64;  
Matches 199; Conservative 107; Mismatches 218; Indels 113; Gaps 11;  
Qy 61 KEILSNTINGMKPG-LNAILGPTGGKSLDLVLAARKDPSGLSGDVLNGAP-RPANFK 118  
Db 1 KEILSGIQGMAHPGCVTAINGASGAGXTTFLDLARKNKGQVSGDFYNGEKVSDPEYK 60  
Qy 119 CNSGYVVQDDVVMGTLTVRENLOFSALRLATTMTNHEKNERNVIOELGLDKVADSKV 178  
Db 61 NAVGFVDQETMLPTLVHETILNSALLRLPKDMTRAAKEORVIEVEKQLGIYHIRDSL 120  
Qy 179 GTQ--FIRGVSGGERKRTSGMELITDPSILFIDEPPTGLDSTA-NAVLLLLKRMKOG 235  
Db 121 GSEEGKRGISGGEKRRVGIACELVTSILFDEPTSGLDAYNAYNVVECLVLTAKYK 180  
Qy 236 RTTIFSTHORYSTFKLFDLSITLLASGLRMLFHGPAQCALGYFESAGVHCEAYNNPADFFL 295  
Db 181 RTVIFTIHPQRSNIVALFDRLILLAQKTYVSGPLHCQYFQIGYTCPPGFNIADYLV 240  
Qy 296 DI-----INGDSTAVALNREDFKA-----TEIIEPS----- 322  
Db 241 DLTWHAGSTSSYDDGTLSDVGVSGPSSTRAVKSIASVSGVSGIDSDSLVSSSSRPENKR 300  
Qy 323 -----KODKPL----- 328  
Db 301 RDSVRRQERELYTRKQAVDTAASSDAGDEIGYKLOKQPPVTPLRSTNDDLDHLPPLA 360  
Qy 329 -----IEKLAIYVNSSFYKETKAELHQL-----SGEKKKKITVFKETSYT----- 370

Db 361 ATGTDLDVLIESYIHSIDIAASTHEEIHQATAAAVNSGQNSGYVADGNI-YTGTMGKY 419  
Qy 371 --TSFCHQLRWVSKRFGKNLGNPQASIAQIVTVVLGLVIGAIYFGLKNDSTGIONRAG 428  
Db 420 ARVGLFRQFVILSQRIWKNLRYRNPMLMTHYATAILLAVFAGLYFGLTLDIAGFQNRLG 479  
Qy 429 VLPFLTTNQCFSSVSAVELFVVEKKLFIHEYISGYRVSVSYFLGKLLSLLPMTLPSII 488  
Db 480 LFFEVLAALFGFSTLTSLGVFSQBRLLFVRBRANGYSPITYFAAKVLFDIVPLRIIPPIL 539  
Qy 489 FTCIVVFMGLGPKADAFVMMETLMMVAYSASSMALATAAGOSVSVATLLMTICFVEM 548  
Db 540 LGALIYPMTGVLVADYQRFVFIIVLVLFNLAAAICLFCILCKDGGVANLIGSLVWLFS 599  
Qy 549 MIFSGLLVNLTTIASLWMLQYFSPRYGFTALOHNE 585  
Db 600 LLFAGLLLNHNNAIPAAALWLQWLSIFHYGFEALIVNE 636

Search completed: March 23, 2004, 18:37:18  
Job time : 59 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2004, 18:28:24 ; Search time 23 Seconds  
(without alignments)  
1470.218 Million cell updates/sec

Title: US-09-961-086-1  
Perfect score: 3352  
Sequence: 1 MSSSNVEVFPVSGQNTNGF.....MIVIFLTAYLKLFLKKYS 655

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/2/1aa/5A-COMB.pep:\*  
2: /cgn2\_6/prodata/2/1aa/5B-COMB.pep:\*  
3: /cgn2\_6/prodata/2/1aa/6A-COMB.pep:\*  
4: /cgn2\_6/prodata/2/1aa/6B-COMB.pep:\*  
5: /cgn2\_6/prodata/2/1aa/PCITUS-COMB.pep:\*  
6: /cgn2\_6/prodata/2/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3352	100.0	655	US-09-245-808-1	Sequence 1, Appli
2	3331	99.4	655	US-09-767-594-1	Sequence 1, Appli
3	567	16.9	617	US-09-614-912-138	Sequence 138, App
4	537	16.0	1296	US-09-614-912-140	Sequence 140, App
5	461	13.8	539	US-09-614-912-144	Sequence 144, App
6	254	7.6	198	US-10-162-012-36	Sequence 36, Appl
7	246	7.3	270	US-09-134-001C-5561	Sequence 5561, Ap
8	240.5	7.2	653	US-09-543-681A-5411	Sequence 5411, Ap
9	238.5	7.1	1684	US-08-665-259-25	Sequence 25, Appl
10	238.5	7.1	1684	US-08-762-500-25	Sequence 25, Appl
11	238.5	7.1	1704	US-08-762-500-75	Sequence 75, Appl
12	232	6.9	256	US-09-134-001C-4600	Sequence 4600, Ap
13	228	6.8	361	US-09-107-532A-6490	Sequence 6490, Ap
14	227.5	6.8	273	US-09-134-001C-3841	Sequence 3841, Ap
15	226.5	6.8	1302	US-08-232-537-2	Sequence 2, Appli
16	224.5	6.7	382	US-09-134-001C-3439	Sequence 3439, Ap
17	222.5	6.6	322	US-09-107-532A-4662	Sequence 4662, Ap
18	222.5	6.6	345	US-09-107-532A-3849	Sequence 3849, Ap
19	222	6.6	1349	US-08-612-734B-2	Sequence 2, Appli
20	220	6.6	229	US-09-134-000C-3584	Sequence 3584, Ap
21	219	6.5	329	US-09-107-532A-4844	Sequence 4844, Ap
22	219	6.5	1457	US-08-665-259-27	Sequence 27, Appl
23	219	6.5	1457	US-08-762-500-27	Sequence 27, Appl
24	218.5	6.5	374	US-09-489-039A-8496	Sequence 8496, Ap
25	217.5	6.5	255	US-09-543-681A-6727	Sequence 6727, Ap
26	216.5	6.5	715	US-09-543-681A-7603	Sequence 7603, Ap
27	215	6.4	240	US-09-489-039A-12478	Sequence 12478, A

ALIGNMENTS

RESULT 1

US-09-245-808-1

; Sequence 1, Application US/9245808

; Patent No. 6313277

; GENERAL INFORMATION:

; APPLICANT: Doyle, L. Austin

; APPLICANT: Abruzzo, Lynne V.

; APPLICANT: Ross, Douglas D.

; TITLE OF INVENTION: Breast Cancer Resistance Protein (BCRP) and DNA which

; FILE REFERENCE: encodes it

; CURRENT APPLICATION NUMBER: US/09/245,808

; CURRENT FILING DATE: 1999-02-05

; EARLIER APPLICATION NUMBER: 60/073763

; EARLIER FILING DATE: 1998-02-05

; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: Patentin Ver. 2.0

; SEQ ID NO 1

; LENGTH: 655

; TYPE: PRT

; ORGANISM: Human MCF-7/AdrVp cells

US-09-245-808-1

Query Match 100.0%; Score 3352; DB 4; Length 655;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 655; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSSSNVEVFPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICVYVKLSGFLPCRKPYE 60

DB 1 MSSSNVEVFPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICVYVKLSGFLPCRKPYE 60

QY 61 KEILSNINGIMKPGNAILGFTGGKSSLDVLAARKDPSGLSGDVLINGAPRANPKCN 120

DB 61 KEILSNINGIMKPGNAILGFTGGKSSLDVLAARKDPSGLSGDVLINGAPRANPKCN 120

QY 121 SGYVQDDVVMGTLTVRENLQFSAALRLATNTNHEKNERINRVQELGLDKVADSKVGT 180

DB 121 SGYVQDDVVMGTLTVRENLQFSAALRLATNTNHEKNERINRVQELGLDKVADSKVGT 180

QY 181 QFIRGVSGGKRKRTSIGMELITDPSILFLDEPTTGLDSSSTANAVLLLLKMSKQGRITIF 240

DB 181 QFIRGVSGGKRKRTSIGMELITDPSILFLDEPTTGLDSSSTANAVLLLLKMSKQGRITIF 240

QY 241 SIHQFRYSIFKLPDLSTLLASGRIMFHGPAQALGYESAGYHCEAYNNPADFLDLING 300

DB 241 SIHQFRYSIFKLPDLSTLLASGRIMFHGPAQALGYESAGYHCEAYNNPADFLDLING 300

QY 301 DSTAVALNREDDFKATEIIEPSKQDKPLIEKLAIEIYVNSSFYKETKAEHLQSLSGGKKK 360

DB 301 DSTAVALNREDDFKATEIIEPSKQDKPLIEKLAIEIYVNSSFYKETKAEHLQSLSGGKKK 360

QY 361 ITVFKISYTTSFCHQLRWVSKRSPKLLGNPQASIAQIIVTVVLGLVIGAIYFGLKND 420  
DB 361 ITVFKISYTTSFCHQLRWVSKRSPKLLGNPQASIAQIIVTVVLGLVIGAIYFGLKND 420  
QY 421 TGIQNRAGVLPFLTTNOCFSSVSAVELFVWEKLFIFHEIISGYRVSSYFGLKLLDLLP 480  
DB 421 TGIQNRAGVLPFLTTNOCFSSVSAVELFVWEKLFIFHEIISGYRVSSYFGLKLLDLLP 480  
QY 481 MTMLPSIIFTCIYVFMGLKPKADAFFVMMFTLMMVAYSASSMALAIAAGOSVSVATLL 540  
DB 481 MTMLPSIIFTCIYVFMGLKPKADAFFVMMFTLMMVAYSASSMALAIAAGOSVSVATLL 540  
QY 541 MTICFVFMIFSGLLVNLTTIASWLSWLOYSIPRYGFTALQHNFLQGNFCPLNATGN 600  
DB 541 MTICFVFMIFSGLLVNLTTIASWLSWLOYSIPRYGFTALQHNFLQGNFCPLNATGN 600  
QY 601 NPCNYATCTGEBYLVKQIDILSPWGLWKNHVALACMIVIFLTIAYLKLLFLKKYS 655  
DB 601 NPCNYATCTGEBYLVKQIDILSPWGLWKNHVALACMIVIFLTIAYLKLLFLKKYS 655

RESULT 2  
US-09-767-594-1  
; Sequence 1, Application US/09767594  
; Patent No. 6521635  
; GENERAL INFORMATION:  
; APPLICANT: Bates, Susan  
; APPLICANT: Robey, Robert  
; APPLICANT: The Government of the United States of America  
; APPLICANT: as represented by the Secretary of the  
; APPLICANT: Department of Health and Human Services  
; TITLE OF INVENTION: Inhibition of MXR Transport by Acridine Derivatives  
; FILE REFERENCE: 015280-402100US  
; CURRENT APPLICATION NUMBER: US/09/767,594  
; CURRENT FILING DATE: 2001-01-22  
; PRIOR APPLICATION NUMBER: US 60/177,410  
; PRIOR FILING DATE: 2000-01-20  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 1  
; TYPE: PRT  
; LENGTH: 655  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: human mitoxanthrone resistance (MXR)/BRCP/ABCP  
US-09-767-594-1

Query Match 99.4%; Score 3331; DB 4; Length 655;  
Best Local Similarity 99.4%; Pred. No. 0;  
Matches 651; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSSNVVEFIPVQNTNGFPATASNDLKAFTEGAVLSFHNICVYRVKLSGFLPCRKPE 60  
DB 1 MSSNVVEFIPVQNTNGFPATVSNLDKAFTEGAVLSFHNICVYRVKLSGFLPCRKPE 60  
QY 61 KEILSNINGIMKPGNALIGPTGGKSLLDVLAARDPSGLSGDVLINGAPRANFKCN 120  
DB 61 KEILSNINGIMKPGNALIGPTGGKSLLDVLAARDPSGLSGDVLINGAPRANFKCN 120  
QY 121 SGYVVQDDVVMGTLTVRENLFSAALRLATTMTNHEKNERINRVIELGLDKVADSKVGT 180  
DB 121 SGYVVQDDVVMGTLTVRENLFSAALRLATTMTNHEKNERINRVIELGLDKVADSKVGT 180  
QY 181 QFIRVSGGERKRTSIGNELITDPSILDEPTTGLDSDSTANAVILLKRMKSGKGRITIF 240  
DB 181 QFIRVSGGERKRTSIGNELITDPSILDEPTTGLDSDSTANAVILLKRMKSGKGRITIF 240  
QY 241 SIHQPRYSIFKLFDSLTLLASGRLMFHGPAQALGYFESAGYHCEAYNNPADFFLDIING 300  
DB 241 SIHQPRYSIFKLFDSLTLLASGRLMFHGPAQALGYFESAGYHCEAYNNPADFFLDIING 300

QY 301 DSTAVALNREEDFKATEIIEPSKQKPLIEKLABIYVNSFYKTKAELHOLSGEKKK 360  
DB 301 DSTAVALNREEDFKATEIIEPSKQKPLIEKLABIYVNSFYKTKAELHOLSGEKKK 360  
QY 361 ITVFKISYTTSFCHQLRWVSKRSPKLLGNPQASIAQIIVTVVLGLVIGAIYFGLKND 420  
DB 361 ITVFKISYTTSFCHQLRWVSKRSPKLLGNPQASIAQIIVTVVLGLVIGAIYFGLKND 420  
QY 421 TGIQNRAGVLPFLTTNOCFSSVSAVELFVWEKLFIFHEIISGYRVSSYFGLKLLDLLP 480  
DB 421 TGIQNRAGVLPFLTTNOCFSSVSAVELFVWEKLFIFHEIISGYRVSSYFGLKLLDLLP 480  
QY 481 MTMLPSIIFTCIYVFMGLKPKADAFFVMMFTLMMVAYSASSMALAIAAGOSVSVATLL 540  
DB 481 MTMLPSIIFTCIYVFMGLKPKADAFFVMMFTLMMVAYSASSMALAIAAGOSVSVATLL 540  
QY 541 MTICFVFMIFSGLLVNLTTIASWLSWLOYSIPRYGFTALQHNFLQGNFCPLNATGN 600  
DB 541 MTICFVFMIFSGLLVNLTTIASWLSWLOYSIPRYGFTALQHNFLQGNFCPLNATGN 600  
QY 601 NPCNYATCTGEBYLVKQIDILSPWGLWKNHVALACMIVIFLTIAYLKLLFLKKYS 655  
DB 601 NPCNYATCTGEBYLVKQIDILSPWGLWKNHVALACMIVIFLTIAYLKLLFLKKYS 655

RESULT 3  
US-09-614-912-138  
; Sequence 138, Application US/09614912  
; Patent No. 6677502  
; GENERAL INFORMATION:  
; APPLICANT: Allen, Steve  
; APPLICANT: Rafalski, Antoni  
; APPLICANT: Orozco, Buddy  
; APPLICANT: Miao, Gou-Hau  
; APPLICANT: Farnodu, Omolayo O.  
; APPLICANT: Lee, Jian Ming  
; APPLICANT: Sakai, Hajime  
; APPLICANT: Weng, Zude  
; APPLICANT: Cai, Perry G  
; APPLICANT: Anderson, Shawn  
; TITLE OF INVENTION: Plant Metabolism Genes  
; FILE REFERENCE: BBI378 US NA  
; CURRENT APPLICATION NUMBER: US/09/614,912  
; CURRENT FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: 60/143,401  
; PRIOR FILING DATE: 1999-07-12  
; PRIOR APPLICATION NUMBER: 60/143,412  
; PRIOR FILING DATE: 1999-07-12  
; PRIOR APPLICATION NUMBER: 60/146,650  
; PRIOR FILING DATE: 1999-07-30  
; PRIOR APPLICATION NUMBER: 60/170,906  
; PRIOR FILING DATE: 1999-12-15  
; PRIOR APPLICATION NUMBER: 60/172,959  
; PRIOR FILING DATE: 1999-12-21  
; PRIOR APPLICATION NUMBER: 60/172,946  
; PRIOR FILING DATE: 1999-12-21  
; NUMBER OF SEQ ID NOS: 204  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 138  
; TYPE: PRT  
; LENGTH: 617  
; ORGANISM: Zea mays  
US-09-614-912-138

Query Match 16.9%; Score 567; DB 4; Length 617;  
Best Local Similarity 28.0%; Pred. No. 8.4e-50;  
Matches 169; Conservative 117; Mismatches 224; Indels 94; Gaps 22;

QY 37 LSFHNCVYRVKLSGFLPCRKPEK-----EILSNINGIMKPG-LNALIGPTGGG 85  
DB 11 MSFDNVNYYVDM-----PAEMKHQGVQDRLQLLREVTGSPRPGVLTALMGVSGAG 61  
QY 86 KSSLDDVLAARDPSGLSGDVLINGAPR-PANFKCNKSGYVQDDVVMGTLTVRENLFQSA 144

Db 62 KTLMDVLGRGTGGYIEGDIIRIAGYKPNQATFARISGYCEQNDIHSQVTVRESLIYA 121  
QY 145 ALRLATMTNHE-----KNERINRVIQELGDKVADSKVGTQFIRGVSGGGERKTSIGME 199  
Db 122 FLRLPGKIGDQBITDDIKQFVDEVMELVDLNDALVGLPGITGLSTEQRKLTIAVE 181  
QY 200 LITDSILEPDTGLDSTANAVLLKRLKMSKQGRITIIHSIHOPRYSIFKLDSLTLL 259  
Db 182 LVANPSIIPMDPBTGLDARAAAIWRTVNTVDTGRTVCTIHOPIIDIFESFDELL 241  
QY 260 -ASGRIMFHGP-----AQEALGYFESA-----GYHCEAYNNPADFFLDIINGSTAVA 306  
Db 242 KRGQVIYSGKLRNSQKWEYFEALPGVPKIKDKY-----NPATWLEV-----SSVATE 292  
QY 307 LNREDFKATEIIEPSKQDKPLIEKLABIYVNSPYKTKAEHLHOLSGEKKKITVPEKE 366  
Db 293 VLKMDP-----AKYETSPLYKQKLVNQLSQPEFGTSDLYP-P 332  
QY 367 ISYTSFCHLRWWSKRSFKNLLGNPQASIAQIITVVVLGVIGAIYFGL---KNDSTGI 423  
Db 333 TYSQSTIQPKACLWKQWLTWRSPDYNLVRSYFTLLVALLGSIWFRIQTNMEDATLL 392  
QY 424 QNRAGVLF-----FLTNQCFSSVSAVELFVVEKKLFIHEYISGYTVRSYFLGKLLSLL 479  
Db 393 GMVIGAMTYAVMFIGINN- -STQPV- -VSIERTVYRERAGYSAMPYAIQVVI-E 448  
QY 480 PMTMLPSIIFTCIVFMLGLKPKADA- - -PFVMMFTLMVAYSASSMALAIAAGQSVVS 535  
Db 449 PIVVQVOTTYTLLVAVMSFQWTAKFVFFISVFSFLYFTY-YGMVAISPNH- -E 504  
QY 536 VATLMTICFVMMIFSGLLVNLTTIASWLSWLOQFSI- - -PRYGTALQHNEFLQON 590  
Db 505 VASIFAAAFSLNLFSGFFIPRIPGW- -WITYVICPLAWTVYGLIVTYQYGLDLELI 562  
QY 591 FCPG 594  
Db 563 SVPG 566

## RESULT 4

US-09-614-912-140  
; Sequence 140, Application US/09614912  
; Patent No. 6677502  
; GENERAL INFORMATION:  
; APPLICANT: Allen, Steve  
; APPLICANT: Rafalski, Antoni  
; APPLICANT: Orozco, Buddy  
; APPLICANT: Miao, Gou-Hau  
; APPLICANT: Famodu, Omolayo O.  
; APPLICANT: Lee, Jian Ming  
; APPLICANT: Sakai, Hajime  
; APPLICANT: Weng, Zude  
; APPLICANT: Caimi, Perry G  
; APPLICANT: Anderson, Shawn  
; TITLE OF INVENTION: Plant Metabolism Genes  
; FILE REFERENCE: BB1378 US NA  
; CURRENT APPLICATION NUMBER: US/09/614,912  
; CURRENT FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: 60/143,401  
; PRIOR FILING DATE: 1999-07-12  
; PRIOR APPLICATION NUMBER: 60/143,412  
; PRIOR FILING DATE: 1999-07-12  
; PRIOR APPLICATION NUMBER: 60/146,650  
; PRIOR FILING DATE: 1999-07-30  
; PRIOR APPLICATION NUMBER: 60/170,906  
; PRIOR FILING DATE: 1999-12-15  
; PRIOR APPLICATION NUMBER: 60/172,959  
; PRIOR FILING DATE: 1999-12-21  
; PRIOR APPLICATION NUMBER: 60/172,946  
; PRIOR FILING DATE: 1999-12-21  
; NUMBER OF SEQ ID NOS: 204  
; SOFTWARE: Microsoft Office 97

; SEQ ID NO 140  
; LENGTH: 1296  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
US-09-614-912-140

Query Match 16.0%; Score 537; DB 4; Length 1296;

Best Local Similarity 27.5%; Pred.No. 3.9e-46; Indels 82; Gaps 22;  
Matches 162; Conservative 115; Mismatches 230;

QY 37 LSFHNICYRYKL-----KSGFLPCRPVKEKEILSNINGIMKPG-LNAIILPTGGKSSLL 90  
Db 690 MSFDDVNYVYDPAEMKQGGWDDR-----LQLLRDVTGSRPAVLTAALGVSGAGKTTLM 745  
QY 91 DVLAAKDDPSGLSDVLIINGAPR-PANFKCNSGVVVODDVVMGTLTVRENLFSAALRLA 149  
Db 746 DVLAGRKTGGYIEGDMRISGYPKQNTFARISGYCEQNDIHSQVTVRESLIYSAFLRP 805  
QY 150 TMTNHE-----KNERINRVIQELGDKVADSKVGTQFIRGVSGGGERKTSIGMELITDP 204  
Db 806 EKIGDQEIITDDIKQFVDEVMELVDLNDKDALVGLPGITGLSTEQRKLTIAVELVANP 865  
QY 205 SILFDEPTGLDSTANAVLLKRLKMSKQGRITIIHSIHOPRYSIFKLDSLTLL-ASGR 263  
Db 866 SIIFMDEPTSGLDARAAAIWRTVNTVDTGRTVCTIHOPIIDIFESFDELLKRGQG 925  
QY 264 LMFHGP-----AQEALGYFESA-----GYHCEAYNNPADFFLDIINGSTAVLNREE 311  
Db 926 VIYSGQLGRNSQKWEYFEALPGVPKIKDKY-----NPATWLEV-----SSVAAEVLNM 976  
QY 312 DFKATEIIEPSKQDKPLIEKLABIYVNSPYKTKAEHLHOLSGEKKKITVPEKEISYTT 371  
Db 977 DF-----AEYKTSPLYKQKLVNQLSQPEFGTSDLYP-PTKYSQ 1016  
QY 372 SFCHLRWWSKRSFKNLLGNPQASIAQIITVVVLGVIGAIYF-GLK-NDSTGIONRAG 428  
Db 1017 STGQFACLWKQWLTWRSPDYNLVRSYFTLLVALLGSIWFRIQTNMEDATLL 1076  
QY 429 VLF-----FLTNQCFSSVSAVELFVVEKKLFIHEYISGYTVRSYFLGKLLSLLPMTML 484  
Db 1077 AMYAVMFIGINN- -ATQPIVSIERTVYRERAGYSAMPYAIQVVI-IPYVFV 1132  
QY 485 PSIIFTCIVFVFMIFSGLLVNLTTIASWLSWLOQFSI- - -YGFALQHNE 585  
Db 1133 QYAYTLLVAVMSFQWTAKFVFFISVFSFLYFTY-YGMVAISPNH- -EVAALF 1188  
QY 541 MTICFVMMIFSGLLVNLTTIASWLSWLOQFSI- - -YGFALQHNE 585  
Db 1189 AAAPYSLEFLNLFSGFFIPRIPGW- -WITYVICPLAWTVYGLIVTYQYGLD 1236

## RESULT 5

US-09-614-912-144  
; Sequence 144, Application US/09614912  
; Patent No. 6677502  
; GENERAL INFORMATION:  
; APPLICANT: Allen, Steve  
; APPLICANT: Rafalski, Antoni  
; APPLICANT: Orozco, Buddy  
; APPLICANT: Miao, Gou-Hau  
; APPLICANT: Famodu, Omolayo O.  
; APPLICANT: Lee, Jian Ming  
; APPLICANT: Sakai, Hajime  
; APPLICANT: Weng, Zude  
; APPLICANT: Caimi, Perry G  
; APPLICANT: Anderson, Shawn  
; TITLE OF INVENTION: Plant Metabolism Genes  
; FILE REFERENCE: BB1378 US NA  
; CURRENT APPLICATION NUMBER: US/09/614,912  
; CURRENT FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: 60/143,401  
; PRIOR FILING DATE: 1999-07-12  
; PRIOR APPLICATION NUMBER: 60/143,412

```
/ PRIOR FILING DATE: 1999-07-12
/ PRIOR APPLICATION NUMBER: 60/146,650
/ PRIOR FILING DATE: 1999-07-30
/ PRIOR APPLICATION NUMBER: 60/170,906
/ PRIOR FILING DATE: 1999-12-15
/ PRIOR APPLICATION NUMBER: 60/172,959
/ PRIOR FILING DATE: 1999-12-21
/ PRIOR APPLICATION NUMBER: 60/172,946
/ PRIOR FILING DATE: 1999-12-21
/ NUMBER OF SEQ ID NOS: 204
/ SOFTWARE: Microsoft Office 97
/ SEQ ID NO 144
/ LENGTH: 539
/ TYPE: PRT
/ ORGANISM: Triticum aestivum
/ FEATURE:
/ NAME/KEY: UNSURE
/ LOCATION: (272)...(273)
US-09-614-912-144

Query Match      13.8%; Score 461; DB 4; Length 539;
Best Local Similarity 25.0%; Pred. No. 7.5e-39;
Matches 148; Conservative 107; Mismatches 241; Indels 96; Gaps 20;

QY 102 LSGDVLINGAP-RPANFKCNSGVVQDDVVMGTLTVRENLOFSALRLATMTNHEKNER 160
Db 4 IEGEITVSGYPKQETFAISGYCEQNDIHSPTHVYESLVFSAWLKLPDAEVDERRKMF 63
QY 161 INRVIOELGLDKVADSKVGTGFIKVGSGGRKRTSIGMELITDPSIILFDEPTTGLDSST 220
Db 64 IEEIMDLVELSLRGALVGLPCVNGLSLSTQKRLTIAVELVANPSIIFMDEPTSGLDARA 123
QY 221 ANAVLLKKNMSKQRTIIFSTHOPRYSIFKLFSLTL-ASGRLMFHPG-----AQBALG 275
Db 124 RAIWRTVNTVNTGRTVCTIHSIDIFAFDELFLMKGGEEIYVGPVQNSANLIE 183
QY 276 YPESA-----GYHCEAYNPADPFLDIINGDSTAVALNREDFKATEIIEPSKDKP 327
Db 184 YFEIEGISKIDGY-----NPATWLE-----VSSSAQEEM 215
QY 328 LIEKLAIEYVNSFFKYEKAEHLQSLGGEKKKTVFKIEISYTSF---CHOLRWVKRS 384
Db 216 LGIDFAEVROSELYORKELIKELSMAPGSSDINF-PTYRSFVTCIACIWKQXXS 274
QY 385 FKNLLGNQASIAQIIVTVGLVIGALYFGLKNDSTGIQN-----RAGVLFFLTN 436
Db 275 YWR---NPSYTAVRLLFIVIALMFGTWFMDLGSKTRESQLEFNAMGMYAAVLYIGVN 331
QY 437 QCFSSVSVELFVTEKKLFIHYSIGYRVSSYFLGKLLSLLPMTLPSIIFTCIVFM 496
Db 332 ----SGSQPVVVVVERTVYERAAAGMYSAPFYAGQVAIE-PPYVLVQALYIGGLVYSM 386
QY 497 LGLKPKADAF----FVMFTLMVAYSASSMALATAAGQSVSVVATILMTICFVFMIFS 552
Db 387 IGFEVTVAKLWLFYFMFTMLYFF-YGMVAVGLTNE---STAALISSAFYVNWLF 442
QY 553 GLLV---NLTTIASLWMLQFSPRYGFTALQHNELFQNGFCPLNATGNPNVATCT 609
Db 443 GYLIPRPKLPVWRWYSICPVANTLYGLVASQFGDI-----QQPLDQGV-P 488
QY 610 GEYLVKQGI-----DLSFWGLKXHVACIMVIFLFIAYLKLLFLKX 653
Db 489 GPQIIVAFVTDYFGFHHDFL-WVVMVHVAVTVLFAELFSFALMRFNQKR 539

RESULT 6
US-10-162-012-36
/ Sequence 36, Application US/10162012
/ Patent No. 6682597
/ GENERAL INFORMATION:
/ APPLICANT: Curtis, Rory A.J.
/ APPLICANT: Silos-Santiago, Inmaculada
/ APPLICANT: Gu, Wei
```



RESULT 7  
 US-09-134-001C-5561  
 ; Sequence 5561, Application US/09134001C  
 ; Patent No. 6380370  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynn Doucette-Stamm et al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: GTC-007  
 ; CURRENT APPLICATION NUMBER: US/09/134,001C  
 ; CURRENT FILING DATE: 1998-08-13  
 ; PRIOR APPLICATION NUMBER: US 60/064,964  
 ; PRIOR FILING DATE: 1997-11-08  
 ; PRIOR APPLICATION NUMBER: US 60/055,779  
 ; PRIOR FILING DATE: 1997-08-14  
 ; NUMBER OF SEQ ID NOS: 5674  
 ; SEQ ID NO 5561  
 ; LENGTH: 270  
 ; TYPE: PRT  
 ; ORGANISM: Staphylococcus epidermidis  
 US-09-134-001C-5561

Query Match 7.3%; Score 246; DB 4; Length 270;  
 Best Local Similarity 27.1%; Pred. No. 66-17;  
 Matches 81; Conservative 56; Mismatches 102; Indels 60; Gaps 12;  
 QY 42 ICYRVKLV-----KSGFLPC-----RKPYKEILSNINGIMKPG-LNALIGPTG 83  
 DB 1 VCLRKIIWKYDRSDYLSLEVRQSKVYGDQAFQVLEKINLDVDEGEFISIMGPG 60  
 QY 84 GKSSLLDVLARKDPGSLGSDVING--APRAN-----FKNSGVVQDDVVMGTLT 135  
 DB 61 SKGTLNLNLSIDYMT--KGSITINGKQLEKLSNKLSDIRKDKIGFIDYINILTLT 118  
 QY 136 VRENTQFSAALRLATMTNHEKNERINRVIOELGLDKVADSKVGTQFIRGVSGBERKRTS 195  
 DB 119 VKENIMPLSVQKLDQKINH---ERYQRIVEALNISDISD-----KYPSELSGGQRQTS 170  
 QY 196 IGNEILTDPSILFLDEPTGLDSTANAVLLLLKRMKQK-RTIIFSIHQPRYSIFKLF 254  
 DB 171 AARAFINLPSIIFADEPTGALDSEKSTLDLLKELKYNBEEFNITLIMVTHDPAASFS--N 228  
 QY 255 SLTLASGLRMLPHGPAQALGFESAGYHCEAYNNPAD---PFLDIINGDSTAVANRE 310  
 DB 229 RVNMLKDGQIF-----TELYQGGDDKQTFYKEILRTQSVLGGINYE 269

RESULT 8  
 US-09-543-681A-5411  
 ; Sequence 5411, Application US/09543681A  
 ; Patent No. 6605709  
 ; GENERAL INFORMATION:  
 ; APPLICANT: GARY BRETON  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
 ; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 2709.1002-001  
 ; CURRENT APPLICATION NUMBER: US/09/543,681A  
 ; CURRENT FILING DATE: 2000-04-05  
 ; PRIOR APPLICATION NUMBER: US 60/128,706  
 ; PRIOR FILING DATE: 1993-04-09  
 ; NUMBER OF SEQ ID NOS: 8344  
 ; SEQ ID NO 5411  
 ; LENGTH: 653  
 ; TYPE: PRT  
 ; ORGANISM: Proteus mirabilis  
 US-09-543-681A-5411

Query Match 7.2%; Score 240.5; DB 4; Length 653;  
 Best Local Similarity 20.1%; Pred. No. 1e-15;  
 Matches 130; Conservative 106; Mismatches 223; Indels 189; Gaps 27;

QY 55 CRKPEKEI--LSNINGIMKPG-----LNAILGPTGGKSSLDVLA 94  
 DB 2 CRRVMSALLENNVSRLYTNGEEDFVLDQLSLTINAGEMVAIIIGASGSGKSTLNNILG 61  
 QY 95 ARKDPGSLGSDVING-----APRANFKNSGVVQDDVVMGTLTVRENLOF 142  
 DB 62 CLDKPS--SGEYKVGQCVADMESDQLAALREHF---GFIFQRYHLMAHLTAEQNVEI 115  
 QY 143 SAALRLATMTNHEKNERINRVIOELGLDKVADSKVGTQFIRGVSGBERKRTSIGMELIT 202  
 DB 116 PA---IYAKSTEQRKERARALLTRGLAERTHYRPSQ-----LSGGQQRVSIARALWN 167  
 QY 203 DPSILFLDEPTGLDSTANAVLLLLKRMKQKRTIIFSIHQPRYSIFKLFPSLTLLASG 262  
 DB 168 GGEVILADEPTGALDSEKSEVMAILKQLNQOQHTVIVTHDP--LIAQQAIRIIEIKDG 225  
 QY 263 RLM-----PHG-----PAQEALGFESA-GVHCEAYNNP----- 290  
 DB 226 QIISDNNNHSAPVKVPFAIQATSYFHOVIGRFTQALNMAWVNVKIRTLTLMGLI 285  
 QY 291 -----ADFFLDIINGDSTAVANREEDFKATEIIBPSKQDK-PLIEKLAETV 337  
 DB 286 IGIASVVTVIIVIGDAAKDRVLADIKAIGANTIDIYPKELGSDSPEDKQSLTIQDVDAK 345  
 QY 338 NSSFYKETKAEHL--QLSGGEKKKKITV-----FKELIS-YTTSFCH---Q 376  
 DB 346 QOSYIOSVTPQIYFSRLRRGNQDAPATVSGVNEGYFSYALKPAQGSTFTPDMIHROAQ 405  
 QY 377 LRWVSKRSFKNLGNPQASIA-QIIVTVVLGLVIGAI-----YFGLKND-----S 420  
 DB 406 VVVIDENTRHRFPFNQAVIGEQLIIRNIPSTIIGVVAEQKSTFGDNKSLRVWVPYSTLS 465  
 QY 421 TGQNRA-----GVLFLT-----TNOCFSSVSAVELFVVE 451  
 DB 466 SRYNRSYLDNITRVKKEGYDASVAPQQLRLTLIRHGKDIFTYNIDSFIAAEKTTOT 525  
 QY 452 KCLF-----IHEYISGYRVSSYFL-----GKLLSDLLPMTMLPSIIFT 490  
 DB 526 MQLFLTIVAVISLVGGIGVMNIMLVSTERTREIGIRMAVGARASDVMMQOPLIESVL-V 584  
 QY 491 CIVYFMGLKPKADAFVVMFTLMVY-----AYSASSMALATAAGOSV 533  
 DB 585 CLVGGLLGI---SLSPAIAFMASMLPNWHFVFPQTALISAFACSTAI 629

RESULT 9  
 US-08-665-259-25  
 ; Sequence 25, Application US/08665259  
 ; Patent No. 6028173  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Landes, Gregory M.  
 ; APPLICANT: Burn, Timothy C.  
 ; APPLICANT: Connors, Timothy D.  
 ; APPLICANT: Dackowski, William R.  
 ; APPLICANT: Van Raay, Terence J.  
 ; APPLICANT: Klinger, Katherine W.  
 ; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,  
 ; NUMBER OF SEQUENCES: 73  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: GENZYME CORPORATION  
 ; STREET: One Mountain Road  
 ; CITY: Framingham  
 ; STATE: Massachusetts  
 ; COUNTRY: United States of America  
 ; ZIP: 01701  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/665,259

FILING DATE: 17-JUN-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Dugan, Deborah A.  
REGISTRATION NUMBER: 37,315  
REFERENCE/DOCKET NUMBER: IGS-9.1  
TELEPHONE: (508) 872-8400  
TELEFAX: (508) 872-5415  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1684 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-665-259-25

Query Match 7.1%; Score 238.5; DB 3; Length 1684;  
Best Local Similarity 20.3%; Pred. No. 8e-15;  
Matches 169; Conservative 116; Mismatches 241; Indels 305; Gaps 39;

QY 66 NINGMKPLGNAIGPTGGKSSLLDVLAARKDPSGLSGDVLING---APRPANFKNSG 122  
DB 533 NLN-LYEQGITVLLGHNGAGKTTLSMLTGLFPPT--SGRAVISGVEISQDMVQIRKSLG 589  
QY 123 YVQDDVVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIOELGLDKVADSKVGTQF 182  
DB 590 LCPQHDIILFDNLTVAHLYFYAQLK---GLSRQKCPPEVKQMLHIGLEKWNRSR----- 641  
QY 183 IRGVSGGKERKTSIGMELITDPSILFDEPTTGLDSSSTANAVLLLLKRMKQGRITIFS 242  
DB 642 SRFLSGGNRRKLSIGIALIAGSKVLLDEPTSGMDAISRAIMDLQOR-QKSDRTIVLT 700  
QY 243 HOPRYSIFK-----LFDLSLTLLASGRLMFHGPA---OEALGYFESAGVHC-----EAYNPN 290  
DB 701 H-----FWDEADLLGDRIAMAKGELQCCSSFLKQYK-----AGHYMTLVKPEHCNP 750  
QY 291 AD-----FFLD----- 296  
DB 751 EDISOLVHHVFNATLESSAGAELSFILPRESTRHEGLFAKLEKKOKELGIASFGASIT 810  
QY 297 -----IING 300  
DB 811 TWEEFLRVKLVDSMDIQAIQIPALQYQHERRASDWDVNSLGCAMDPSDGIKALIEE 870  
QY 301 DSTAVLN-----REDFKATEIIEP-----SKQ- 324  
DB 871 ERTAVKLTGLALHQQFWMFLKKAYSREWEMVAQVLVPLTCTVTLALLAINYSSEL 930  
QY 325 -DKPLI-----EKLA-----IYVNSFFYKTKAEHLQ----- 351  
DB 931 FDPMLRLTLGEGYGTVPVFPVPGTSQGLQOLSEHLKDALQAEQEPREVLGLDEEFLIF 990  
QY 352 ---LSGGEKKKITVFEKISYTTSPCHLRWVSKRSFKNLGNPQ-----ASIAQI 399  
DB 991 RASVEGGGFNERCLV-----AASF-----RDVGERTVVALFNQAVHSPATALAVVDNL 1040  
QY 400 IVTVVLGLVIGAIYGLKNDSTGIGQ-----NRAGVLFLTNTQCF-----SUSAV 445  
DB 1041 LFKLLCGPHASIVVSNFPQPSALQAKDQFNEGRKGFIDIALNLLFAMAFIATFSTILAV 1100  
QY 446 ELFVVEKKLFIH-EYISGYRVSSYFLGKLLSDLLPMTMLPSIIFTCIVVFMGLKPEKAD 504  
DB 1101 SERAVQAK---HVQFVSGVH-VASFVLSALLWDLISF-LIFSLLL-LVVF-----X 1145  
QY 505 AFFVMFLMVAAYASASSWALIAAGQSVSVATLLMTICVFMFISGLLVNUTTIASW 564  
DB 1146 ADVRAFT--RDGHVADTLLLLGLGVAIIPLMYLMN-----PFLGAATAYTLETTIENI 1198  
QY 565 LSWLOYF-----SIPRYGF-----TALQNEFLGONFCPGLNATCNPNVAT---CTGE 611  
DB 1199 LSGIATFLMVTIMRIPAVKLEBSLKTLDHVLVLPNHCLGM-AVSSFVENYETRYCTSS 1257

QY 612 E-----YLVKQIDILSPW---GLMKHVALACMIVIFLITAYLKLFL 651  
DB 1258 EVAHYCKKNIQOENFYAWSAPGVGRFVMAAS-----GCAYLILFL 1303

RESULT 10  
US-08-762-500-25  
Sequence 25, Application US/08762500  
Patent No. 6030806  
GENERAL INFORMATION:  
APPLICANT: Landes, Gregory M.  
APPLICANT: Burn, Timothy C.  
APPLICANT: Connors, Timothy D.  
APPLICANT: Dackowski, William R.  
APPLICANT: Van Raay, Terence J.  
APPLICANT: Klinger, Katherine W.  
TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,  
TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME  
NUMBER OF SEQUENCES: 83  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENZYME CORPORATION  
STREET: One Mountain Road  
CITY: Framingham  
STATE: Massachusetts  
COUNTRY: United States of America  
ZIP: 01701  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/762,500  
FILING DATE: 09-DEC-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/665,259  
FILING DATE: 17-JUN-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/10469  
FILING DATE: 17-JUN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Dugan, Deborah A.  
REGISTRATION NUMBER: 37,315  
REFERENCE/DOCKET NUMBER: IGS-9.3  
TELEPHONE: (508) 872-8400  
TELEFAX: (508) 872-5415  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1684 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-762-500-25

Query Match 7.1%; Score 238.5; DB 3; Length 1684;  
Best Local Similarity 20.3%; Pred. No. 8e-15;  
Matches 169; Conservative 116; Mismatches 241; Indels 305; Gaps 39;

QY 66 NINGMKPLGNAIGPTGGKSSLLDVLAARKDPSGLSGDVLING---APRPANFKNSG 122  
DB 533 NLN-LYEQGITVLLGHNGAGKTTLSMLTGLFPPT--SGRAVISGVEISQDMVQIRKSLG 589  
QY 123 YVQDDVVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIOELGLDKVADSKVGTQF 182  
DB 590 LCPQHDIILFDNLTVAHLYFYAQLK---GLSRQKCPPEVKQMLHIGLEKWNRSR----- 641  
QY 183 IRGVSGGKERKTSIGMELITDPSILFDEPTTGLDSSSTANAVLLLLKRMKQGRITIFS 242  
DB 642 SRFLSGGNRRKLSIGIALIAGSKVLLDEPTSGMDAISRAIMDLQOR-QKSDRTIVLT 700  
QY 243 HOPRYSIFK-----LFDLSLTLLASGRLMFHGPA---OEALGYFESAGVHC-----EAYNPN 290

Db 701 H-----FDEADLLGDRATMAKGBELQCGSSFLKQKYG-----AGYHMTLVKPEHCNP 750  
QY 291 AD-----FFLD-----296  
Db 751 EDISQLVHHVNPATLESSSAGAEBSFILPRESTRPEGLFAKLEKKQKELGSIASFGASIT 810  
QY 297 -----IING 300  
Db 811 TMEVFLRVKGLVDSSMDIQAIQIPALQYQHERRASDAVDSNLCGAMPDPSDIGALIEE 870  
QY 301 DSTAVALN-----REEDFKATEIIEP-----SKQ- 324  
Db 871 ERTAVKLTGLALHCOQFWAMFLKKAAYSWEKRWAAQVLPVLTCTVTALLAINYSSEL 930  
QY 325 -DKPLI-----EKLAEE-----IYVNSFFKETEKAELHQ-----351  
Db 931 FDDPMLRLTIGEYGRVTVVPSVPGTSQGLQQLSEHLKDALQAEQEPREVGLDLEFLIF 990  
QY 352 ---LSGEGKKKKTIVFKEISYTTSPCHQLRWKSKRSFKNLLGNPQ-----ASIAQI 399  
Db 991 RASVEGGGFNERCLV-----AASF-----RDVGERTVNNALFNNQAYHSPATALAVVDNL 1040  
QY 400 IVTVVLGLVIGAYFGLKNDSTGIQ-----NRAGVLFFLTNNQCFSS-----SVSAV 445  
Db 1041 LFKLLCGPHASIVVSNFPQPSRSLQAQKQDFNEGRKGFIALNLLFAMAFLASTFSLAV 1100  
QY 446 ELFVVEKKKLFTH-EYISGYRVSSYFLGKLLSDLLPMTMLPSIIFTCIVVFMGLKPKAD 504  
Db 1101 SERAVQAK--HVQFVSGVH-VASFWSALLWDLISF-LIPSLLL--LVVF-----X 1145  
QY 505 AFFVMMFTLMMVAYSASSMALATAAGOSVVSATLMTICFVFMIFSGLLVNLTIASW 564  
Db 1146 AFDVRAFT--RDGHMADTLLLLLYGWAILPMLN-----FFFLGATATVTRLTIFI 1198  
QY 565 LSWLQYF-----SIPRYGF-----TALQNEFLGQFCPLNATGNPCNYAT--CTGE 611  
Db 1199 LSGIATFLMTIMRIPAVKLEELSKTLDVFLVLPNHLGM-AVSGFYENYETRYCTSS 1257  
QY 612 E-----YLVKQGDLSFN-----GLWKNHVALACMIVFLFIAYLKLF 651  
Db 1258 EVAHYCKYNIQYQENFYANSAPGVGRFVSAAS-----GCAYLILULF 1303

## RESULT 11

US-08-762-500-75  
; Sequence 75, Application US/08762500  
; Patent No. 6030806  
; GENERAL INFORMATION:  
; APPLICANT: Landes, Gregory M.  
; APPLICANT: Burn, Timothy C.  
; APPLICANT: Connors, Timothy D.  
; APPLICANT: Dackowski, William R.  
; APPLICANT: Van Raay, Terence J.  
; APPLICANT: Klinger, Katherine W.  
; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,  
; TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME  
; NUMBER OF SEQUENCES: 83  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: GENZYME CORPORATION  
; STREET: One Mountain Road  
; CITY: Framingham  
; STATE: Massachusetts  
; COUNTRY: United States of America  
; ZIP: 01701  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/762,500  
; FILING DATE: 09-DEC-1996

CLASSIFICATION: 435  
PRIOR APPLICATION DATA: US 08/665,259  
FILING DATE: 17-JUN-1996  
PRIOR APPLICATION DATA: PCT/US96/10469  
FILING DATE: 17-JUN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Dugan, Deborah A.  
REGISTRATION NUMBER: 37,315  
REFERENCE/DOCKET NUMBER: IG5-9.3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (508) 872-8400  
TELEFAX: (508) 872-5415  
INFORMATION FOR SEQ ID NO: 75:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1704 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-762-500-75  
Query Match 7.1%; Score 238.5; DB 3; Length 1704;  
Best Local Similarity 20.3%; Pred. No. 8.2e-15;  
Matches 169; Conservative 116; Mismatches 241; Indels 305; Gaps 39;  
QY 66 NINIMKPGNLAILGPTGGGKSSLLDVLAKRDPGSLGSDVLING---APRPANFKCSG 122  
Db 553 NLN-LYEGQITVLLGHNGAGKTTLSMLTGLFPPT--SGRAYISGYEISQDMVQIRKSLG 609  
QY 123 YVODDVMVGTLVRENLFQSAALRATTTNHEKNERINRVIOELGLDKVADSKVGTQF 182  
Db 610 LCPQHDILFNLVVAEHLFYAQLK---GLSRQKCPVEVKOMHLIGLEKWNRS-----661  
QY 183 IRGVSGERKRTSGMELITDPSILFDEPTTGLDSTANAVALLLKRMSKQGRITFPI 242  
Db 662 SRFLSGGWRKRLSIGIALIAGSKVLIIDPTSGMDAISRRAIWDLLQR-QKSDRTIIVLT 720  
QY 243 HQPRYSIFK---LPDSLTLASGLRMFGPA---QEALGYFESAGYHC---EAYNPP 290  
Db 721 H-----FMDEADLLGDRATMAKGBELQCGSSFLKQKYG-----AGYHMTLVKPEHCNP 770  
QY 291 AD-----FFLD-----296  
Db 771 EDISQLVHHVNPATLESSSAGAEBSFILPRESTRPEGLFAKLEKKQKELGSIASFGASIT 830  
QY 297 -----IING 300  
Db 831 TMEVFLRVKGLVDSSMDIQAIQIPALQYQHERRASDAVDSNLCGAMPDPSDIGALIEE 890  
QY 301 DSTAVALN-----REEDFKATEIIEP-----SKQ- 324  
Db 891 ERTAVKLTGLALHCOQFWAMFLKKAAYSWEKRWAAQVLPVLTCTVTALLAINYSSEL 950  
QY 325 -DKPLI-----EKLAEE-----IYVNSFFKETEKAELHQ-----351  
Db 951 FDDPMLRLTIGEYGRVTVVPSVPGTSQGLQQLSEHLKDALQAEQEPREVGLDLEFLIF 1010  
QY 352 ---LSGEGKKKKTIVFKEISYTTSPCHQLRWKSKRSFKNLLGNPQ-----ASIAQI 399  
Db 1011 RASVEGGGFNERCLV-----AASF-----RDVGERTVNNALFNNQAYHSPATALAVVDNL 1060  
QY 400 IVTVVLGLVIGAYFGLKNDSTGIQ-----NRAGVLFFLTNNQCFSS-----SVSAV 445  
Db 1061 LFKLLCGPHASIVVSNFPQPSRSLQAQKQDFNEGRKGFIALNLLFAMAFLASTFSLAV 1120  
QY 446 ELFVVEKKKLFTH-EYISGYRVSSYFLGKLLSDLLPMTMLPSIIFTCIVVFMGLKPKAD 504  
Db 1121 SERAVQAK--HVQFVSGVH-VASFWSALLWDLISF-LIPSLLL--LVVF-----X 1165  
QY 505 AFFVMMFTLMMVAYSASSMALATAAGOSVVSATLMTICFVFMIFSGLLVNLTIASW 564  
Db 1166 AFDVRAFT--RDGHMADTLLLLLYGWAILPMLN-----FFFLGATATVTRLTIFI 1218



;; PRIOR APPLICATION NUMBER: US 60/055,779  
;; PRIOR FILING DATE: 1997-08-14  
;; NUMBER OF SEQ ID NOS: 5674  
;; SEQ ID NO 3641  
;; LENGTH: 273  
;; TYPE: PR1  
;; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-3641

Query Match 6.8%; Score 227.5; DB 4; Length 273;  
Best Local Similarity 29.0%; Pred. No. 5.2e-15;  
Matches 76; Conservative 45; Mismatches 78; Indels 63; Gaps 13;  
QY 61 KEILSININGMKPG-LNAILGPTGGKSLDLVLAARKDPSGL-----SGDVLINGAPRPA 115  
DB 31 KHVLENINIKINGEFLAIVPGNAGKSTLLKVL-----GLLPQKGEIIVDGKP--- 81  
QY 116 NFKNS-----GTVQ-----DDVMGTLTVRENJ--QPSAALRLATMTNH 155  
DB 82 -FKGKSSLSKISVSQKAFNAGFPASVKEVLSGLTKTKLQRF-----NK 129  
QY 156 EKNERINRVQELGLDKVADSKVGTQFIRGVSGGERKRTSIGMELITDPSILFLDEPTTG 215  
DB 130 NDQKVTKVLERNISHLINKNAE-----LSGGQOORVLIARALISDSVLVDEPTNG 184  
QY 216 LDSTANAVALLKRMKQRTIIFSIHQPRYSIFKLFDSLTLA--SGRLMHPGPAQ-- 271  
DB 185 IDAKHVSIFYDLDKLGKQGITIILVTH-----DIGVVADTATEVACLKHLHFGHSTAP 240  
QY 272 EALGYEPESAGYHCEAYNNPADF 293  
DB 241 KSLDEVEIS-----KIYGHPIQF 258

## RESULT 15

US-08-232-537-2  
; Sequence 2, Application US/08232537  
; Patent No. 5516655  
; GENERAL INFORMATION:  
; APPLICANT: Peery, Robert B.  
; APPLICANT: Skatrud, Paul L.  
; TITLE OF INVENTION: MULTIPLE DRUG RESISTANCE GENE OF  
; TITLE OF INVENTION: AUREOBASIDIUM PULLULANS  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Eli Lilly and Company  
; STREET: Lilly Corporate Center  
; CITY: Indianapolis  
; STATE: Indiana  
; COUNTRY: USA  
; ZIP: 46285  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,537  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Plant, Thomas G.  
; REGISTRATION NUMBER: 35,784  
; REFERENCE/DOCKET NUMBER: X9212  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 317-276-2459  
; TELEFAX: 317-276-1917  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1302 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein

## US-08-232-537-2

Query Match 6.8%; Score 226.5; DB 1; Length 1302;  
Best Local Similarity 20.4%; Pred. No. 9.2e-14;  
Matches 139; Conservative 102; Mismatches 231; Indels 209; Gaps 27;  
QY 4 SNVEVFPVSGNTNGFPATASNDLKAFTEGAVLSFHNICRYVKLSGFLPCRKPVEKEI 63  
DB 383 SEINVSAGEQEA-----ESDMK-----ADIVFENVTF-----VYPAR--TSARA 421  
QY 64 LSNINGMKPG-LNAILGPTGGKSLDLVLAARKDPSGLSGDVLINGAPRPAFKNS- 121  
DB 422 LEEMSLIKAGOMNAIVGTSGCGKSTLVSLRLYDIS--SGQLTI--GSHDIKDFNVRSL 478  
QY 122 --GVYQDDVMGTLTVRENLOPSAALRLATMTWHEKNERINRVQELGLDKV----- 173  
DB 479 RXYTALVDQSVLPSGVLNLSYGLG---EHSLSDDVLERCTEAKAANLQFVDFLPQ 535  
QY 174 -ADSKVGTQFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVALLKRM 232  
DB 536 GIHTRIGNGGYTSLSGQGNQRICLARALVKKPAALLLDEPTAALDANSEGLINDAVKSA 595  
QY 233 KQRTIIFSIHQPRYSIFKLFDSLTLASGRLMFHPGPAQAL---GYFESAGYHCEAYN 288  
DB 596 ATGTTVMVAH--RLSTVSDSPNIVLMGAKVIEQGNHDELMQLEGAYFNL--IQAOQLN 651  
QY 289 NPADFIDDIINGDSTAVA---LNREDFKA---TEIIEP--SKQDKPLIEKLABIYVNSS 340  
DB 652 DADESSAABVSAATTSQVTPQKASKSDSAASSDTETVPPQAKKEDKPA----- 699  
QY 341 FYKETKAELHOLSGEKKKTIYFKEISYTTSPCHQLRWVSKGSFKNLGNPQASIAQII 400  
DB 700 -----KKAGFWKLLRLCLRLAKSDSPIALG-----LAASI 730  
QY 401 VTVVLGLVIG-AIYFGLKNDSTGIQNEAGVLFELTTNQCFSSVSASVELFVVEKKLFIHEY 459  
DB 731 VSG--GIILGEAIVFG-----NLISVNLDESPE----- 757  
QY 460 ISGYRVSSYFLGKLLSLLPMTMLPSIITCIYVFMGLKPKKADAFVMMFLMVA-- 517  
DB 758 -----FRSRADLPSLFFILALIALF 778  
QY 518 -YSASSMALAIAGQSVSVVATL-LMTICFVFMIFSG-----LLVNLTTIASWLSWLO 570  
DB 779 SYAGNGCCFGIVSSHFAKIQHISLASILRQDMQWFSQSQSVPSLSSSDAGQLCLSG 838  
QY 571 PSIPRYGFTALQHNFEFLQNPCLNATGNPNYATCTGEEYLVKQIDLSFPGWLKXNH 630  
DB 839 VAI-----GTIFTVCVSTIG-----GIILAHVVAWKIA 866  
QY 631 VALACMIVIFLTIAYLKLLFL 651  
DB 867 VLLAAVPMITAGYVRLVL 887

Search completed: March 23, 2004, 18:32:19  
Job time : 33 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2004, 18:23:18 ; Search time 60 Seconds  
(without alignments)  
3084.477 Million cell updates/sec

Title: US-09-961-086-1  
Perfect score: 3352  
Sequence: 1 MSSNVVEFIPVSGQTNGF.....MIVIFLTIAYLKLLFLKKYS 655

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues  
Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*

- 1: Geneseq1980s:\*
- 2: Geneseq1990s:\*
- 3: Geneseq2000s:\*
- 4: Geneseq2001s:\*
- 5: Geneseq2002s:\*
- 6: Geneseq2003as:\*
- 7: Geneseq2003bs:\*
- 8: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3352	100.0	655	5 AAU80029	Aau80029 Human ABC
2	3352	100.0	655	2 AAY5221	Aay5221 Breast Ca
3	3346	99.8	655	4 AAB60104	Aab60104 Human tra
4	3346	99.8	655	5 AAO14781	Aao14781 Human BCR
5	3346	99.8	655	5 AAU80028	Aau80028 Human ABC
6	3346	99.8	655	6 ABR58077	Abr58077 Human ABC
7	3346	99.8	655	6 ADA10917	Ada10917 Human cdn
8	3346	99.8	655	7 ADC54182	Adc54182 Human bre
9	3339	99.6	655	5 AAO14783	Aao14783 Human BCR
10	3338	99.6	655	5 ABB07273	Abb07273 Human BCR
11	3331	99.4	655	3 AAY95365	Aay95365 ATP-bindl
12	3331	99.4	655	4 AAU04348	Aau04348 Human BCR
13	3331	99.4	655	5 ABB52127	Abp52127 Homo sapi
14	3331	99.4	655	5 ABB07270	Abb07270 Human BCR
15	3331	99.4	655	6 ABU63376	Abu63376 Human mit
16	3331	99.4	655	5 AAO14782	Aao14782 Human BCR
17	3053.5	91.1	604	2 AAU73627	Aaw73627 Human sec
18	3053.5	91.1	604	5 ABB61858	Abp61858 Human pol
19	2757	82.2	657	5 ABB07272	Abb07272 Murine BC
20	2325	69.4	456	4 AAB93564	Aab93564 Human pro
21	812	24.2	687	4 ABB59384	Abb59384 Drosophil
22	808	24.1	687	3 AAY78981	Aay78981 Silkworm
23	793.5	23.4	625	3 AAG18080	Aag18080 Arabidops
24	783.5	23.4	632	3 AAG18079	Aag18079 Arabidops
25	783.5	23.4	648	3 AAG18078	Aag18078 Arabidops

26	746	22.3	649	3 AAG43100	Aag43100 Arabidops
27	746	22.3	653	3 AAG43099	Aag43099 Arabidops
28	727	21.7	832	4 ABB59544	Abb59544 Drosophil
29	720	21.5	616	3 AAG50555	Aag50555 Arabidops
30	714.5	21.3	666	5 ABB57112	Abb57112 Mouse isc
31	711	21.2	646	5 AAE28968	Aae28968 Human ABC
32	710.5	21.2	674	5 ABB52126	Abp52126 Homo sapi
33	710	21.2	638	5 ABB98349	Abb98349 Human ABC
34	710	21.2	646	5 AAE28964	Aae28964 Human ABC
35	710	21.2	646	5 ABB98348	Abb98348 Human ABC
36	708	21.1	627	6 ABO07272	Abp07272 Human p53
37	708	21.1	646	6 ABB99694	Abb99694 Amino aci
38	707	21.1	646	5 AAO14186	Aao14186 Human tra
39	706.5	21.1	662	6 ABO07271	Abp07271 Human p53
40	702.5	21.0	663	6 ABB82647	Abb82647 Human Dev
41	700	20.9	646	6 ABB99696	Abb99696 Amino aci
42	694.5	20.7	652	5 AAU96985	Aau96985 Mouse ABC
43	693.5	20.7	652	5 AAE13308	Aae13308 Mouse sit
44	693.5	20.7	652	5 AAE13289	Aae13289 Mouse sit
45	693.5	20.7	652	6 AAE31702	Aae31702 Mouse ABC

ALIGNMENTS

RESULT 1  
AAU80029  
ID AAU80029 standard; protein; 655 AA.  
AC AAU80029;  
XX  
DT 15-JUL-2002 (first entry)  
XX  
DE Human ABCG2 mutant 482T.  
XX  
KW Human; ABCG2; transporter protein; anticancer drug tolerance;  
KW indocarbazole; mutant; muten.  
XX  
OS Homo sapiens.  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 482  
FT /note= "Wild type Arg substituted by Thr"  
XX  
FN WO200228894-A1.  
XX  
PD 11-APR-2002.  
XX  
PF 18-SEP-2001; 2001WO-JP008112.  
XX  
PR 03-OCT-2000; 2000JP-00303441.  
XX  
PA (BANY ) BANYU PHARM CO LTD.  
XX  
PI Komatani H, Hara Y, Kotani H, Nakagawa R;  
XX  
DR WPI: 2002-352228/38.  
DR N-PSDB; ABK49911.  
XX  
PT ABCG2 gene encoding transporter protein capable of selectively  
PT transporting indocarbazole compounds, useful in screening inhibitors and  
PT anticancer agents for administration in chemotherapy.  
XX  
PS Disclosure; Page 87-90; 98pp; Japanese.  
XX  
CC The invention relates to an ABCG2 gene encoding a transporter protein  
CC capable of imparting tolerance to an anticancer agent in mammals  
CC comprising a fully defined sequence as given in the specification or an  
CC amino acid sequence based on the sequence but with some amino acids  
CC substituted, deleted or added. The gene and encoded protein are useful in  
CC screening inhibitors and anticancer agents for administration in  
CC chemotherapy with enhancement in sensitivity of cancer cell tolerance.

CC The gene relating to drug tolerance can be modified e.g. with the  
 CC transporter inhibitors, screened compounds, antibodies and antisense  
 CC nucleosides. The transporter is capable of selectively transporting  
 CC nucleoside compounds extracellularly. The present sequence represents  
 CC the amino acid sequence of human ABCG2 mutant 492T  
 XX  
 SQ Sequence 655 AA;

Query Match 100.0%; Score 3352; DB 5; Length 655;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 555; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSSNNVEVPIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPYE 60  
 Db 1 MSSNNVEVPIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPYE 60

QY 61 KEILSNINGIMKPGNALIGPTGGKSSLLDLAARKDPSGLSDVLIINGAPRANFKCN 120  
 Db 61 KEILSNINGIMKPGNALIGPTGGKSSLLDLAARKDPSGLSDVLIINGAPRANFKCN 120

QY 121 SGYVQDDVVMGTLTVRENLFQSAALRLATTMTNHEKNERINRVIOELGLDKVADSKVGT 180  
 Db 121 SGYVQDDVVMGTLTVRENLFQSAALRLATTMTNHEKNERINRVIOELGLDKVADSKVGT 180

QY 181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVLLLLKMSKQGRITIF 240  
 Db 181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVLLLLKMSKQGRITIF 240

QY 241 SIHQPRYSIFKLFDSLTLASGRMLFHGPAQEAALGYFESAGYHCEAYNNPADFLDIING 300  
 Db 241 SIHQPRYSIFKLFDSLTLASGRMLFHGPAQEAALGYFESAGYHCEAYNNPADFLDIING 300

QY 301 DSTAVANREEDFKATEIIEPSKQDKPLIEKLAETVYNSFYKTKAELHQLSGEKKKX 360  
 Db 301 DSTAVANREEDFKATEIIEPSKQDKPLIEKLAETVYNSFYKTKAELHQLSGEKKKX 360

QY 361 ITVFKESVYTSFCHQLRWKSKRPNLLGNPQASIAQIIVTVVLGLVIGAIYFGLKNDK 420  
 Db 361 ITVFKESVYTSFCHQLRWKSKRPNLLGNPQASIAQIIVTVVLGLVIGAIYFGLKNDK 420

QY 421 TGIONRAGVLPFLTTNQCFSSVSAVELFVVEKLFIEHYISGYRVSSYFGLKLLDLP 480  
 Db 421 TGIONRAGVLPFLTTNQCFSSVSAVELFVVEKLFIEHYISGYRVSSYFGLKLLDLP 480

QY 481 MTMLPSIIPTCIVYFMLGLKPRADAFVVMFTLMWVAYSASSMALAIAGQSVSVATLL 540  
 Db 481 MTMLPSIIPTCIVYFMLGLKPRADAFVVMFTLMWVAYSASSMALAIAGQSVSVATLL 540

QY 541 MTICFVFMIFGILLVNLTTIASWLSWLOYSIPRYGFTALQHNFLGQFCPGLNATGN 600  
 Db 541 MTICFVFMIFGILLVNLTTIASWLSWLOYSIPRYGFTALQHNFLGQFCPGLNATGN 600

QY 601 NPNVATCTGEEYLVKQGDIDSPWGLKXNHVALACMIVIFLTIAVLKLLFLKKYS 655  
 Db 601 NPNVATCTGEEYLVKQGDIDSPWGLKXNHVALACMIVIFLTIAVLKLLFLKKYS 655

RESULT 2  
 AAY15221  
 ID AAY15221 standard; protein; 663 AA.  
 XX  
 AC AAY15221;  
 XX  
 XX 09-NOV-1999 (first entry)  
 DT  
 XX Breast Cancer Resistance Protein (BCRP).  
 DE  
 XX breast cancer; drug resistance; ATP-binding cassette; ABC;  
 KW xenobiotic transporter; chemotherapy; mitoxantrone; doxorubicin;  
 KW breast cancer resistance protein; BCRP.  
 XX  
 OS Homo sapiens.  
 XX

Key Location/Qualifiers  
 FH Domain 87..95 /note= "Walker A motif"  
 FT Domain 221..236 /note= "Phosphopantetheine site"  
 FT Modified-site 345..347 /note= "Glycosylation site on N"  
 FT Region 405..422 /label= TM1  
 FT Modified-site 425..427 /note= "Transmembrane region"  
 FT Region 546..563 /label= TM2  
 FT Modified-site 564..566 /note= "Glycosylation site on N"  
 FT Modified-site 604..606 /note= "Glycosylation site on N"  
 FT Region 638..655 /label= TM3  
 XX WO9940110-A1.  
 PN 12-AUG-1999.  
 XX 05-FEB-1999; 99WO-US002577.  
 XX 05-FEB-1998; 98US-0073763P.  
 XX (UYMA-) UNIV MARYLAND BALTIMORE.  
 XX Ross DD, Doyle LA, Abruzzo L;  
 WPI: 1999-494273/41.  
 N-PSDB; AA206360.  
 XX New breast cancer resistance protein useful for production of antibodies  
 to inhibit resistance activity for enhancing chemotherapy treatment.  
 Claim 4; Fig 2a; 80pp; English.  
 XX The Breast Cancer Resistance Protein (BCRP) is an ATP-binding cassette  
 (ABC) transporter protein. It has a molecular mass of approximately 72.3  
 kilodaltons (kD) exclusive of any glycosylation. Expression of BCRP in  
 drug sensitive human cancer cells confers resistance to mitoxantrone, in  
 doxorubicin, and daunorubicin, and reduces daunorubicin accumulation in  
 the cloned transfected cells. The protein is useful for producing  
 antibodies and antisense probes, which can be used to inhibit the  
 activity of BCRP, therefore enhancing a cancer patient's chemotherapy  
 treatment. The antibodies and probes overcome the problems of breast  
 cancer resistance proteins to make chemotherapy treatment more effective  
 XX  
 SQ Sequence 663 AA;

Query Match 100.0%; Score 3352; DB 2; Length 663;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 655; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSSNNVEVPIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPYE 60  
 Db 9 MSSNNVEVPIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPYE 68

QY 61 KEILSNINGIMKPGNALIGPTGGKSSLLDLAARKDPSGLSDVLIINGAPRANFKCN 120  
 Db 69 KEILSNINGIMKPGNALIGPTGGKSSLLDLAARKDPSGLSDVLIINGAPRANFKCN 128

QY 121 SGYVQDDVVMGTLTVRENLFQSAALRLATTMTNHEKNERINRVIOELGLDKVADSKVGT 180  
 Db 129 SGYVQDDVVMGTLTVRENLFQSAALRLATTMTNHEKNERINRVIOELGLDKVADSKVGT 188

QY 181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVLLLLKMSKQGRITIF 240  
 Db 189 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVLLLLKMSKQGRITIF 248

CC the diagnosis and treatment of transport, metabolic, neurological,  
CC reproductive, cardiovascular and immune disorders, and cell proliferative  
CC disorders such as cancer  
XX  
SQ Sequence 655 AA;

Query Match 99.8%; Score 3346; DB 4; Length 655;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 654; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MSSSNEVFIPVSGNTGPPATASNDLKAFTEGAVLSFHNICIRYVKLSGFFPCRPVE 60  
Db 1 MSSSNEVFIPVSGNTGPPATASNDLKAFTEGAVLSFHNICIRYVKLSGFFPCRPVE 60  
Qy 61 KEILSNINGIMKGLNAIIGPTGGKSSLLDVAARKDPSGLSGDVLINGAPPANFKCN 120  
Db 61 KEILSNINGIMKGLNAIIGPTGGKSSLLDVAARKDPSGLSGDVLINGAPPANFKCN 120  
Qy 121 SGYVQDDVVMGTLTVRENIQFSAALRLATMTNHEKNERINRVIQELGLDKVADSKVGT 180  
Db 121 SGYVQDDVVMGTLTVRENIQFSAALRLATMTNHEKNERINRVIQELGLDKVADSKVGT 180  
Qy 181 QFIRGVSGERKRTSGMELITDPSILFLDEPTTGLDSTANAVALLKEMSKOGRITIF 240  
Db 181 QFIRGVSGERKRTSGMELITDPSILFLDEPTTGLDSTANAVALLKEMSKOGRITIF 240  
Qy 241 SIHQPRYSIFKLFDSLTLASGRIMFHPGPAQEAALGYFESAGYHCEAYNNPADFFLDIING 300  
Db 241 SIHQPRYSIFKLFDSLTLASGRIMFHPGPAQEAALGYFESAGYHCEAYNNPADFFLDIING 300  
Qy 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIEIYVNSSFYKETKAEHLQSGGEKKK 360  
Db 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIEIYVNSSFYKETKAEHLQSGGEKKK 360  
Qy 361 ITVFEKISYTTSFCHQLRWVSQRSFKNLLGNPQASIAQIITVTVGLVIGAIYFGLKXDS 420  
Db 361 ITVFEKISYTTSFCHQLRWVSQRSFKNLLGNPQASIAQIITVTVGLVIGAIYFGLKXDS 420  
Qy 421 TGIQNRAGVLFFLTNNQCFSSVAVELFVVEKKLFHEIYISGYRVSSYFLGKLLSDLLP 480  
Db 421 TGIQNRAGVLFFLTNNQCFSSVAVELFVVEKKLFHEIYISGYRVSSYFLGKLLSDLLP 480  
Qy 481 MTMLPSIIFTCIVVFMGLKPKADAFVVMFTLMVAYSASSMALAIAAGQSVSVATLL 540  
Db 481 MTMLPSIIFTCIVVFMGLKPKADAFVVMFTLMVAYSASSMALAIAAGQSVSVATLL 540  
Qy 541 MTICFVFMIFPSGLLVNLTITIASWLSWLYQFSIPRYGFTALQHNBEFLGQFCPLNATGN 600  
Db 541 MTICFVFMIFPSGLLVNLTITIASWLSWLYQFSIPRYGFTALQHNBEFLGQFCPLNATGN 600  
Qy 601 NPCNYATCTGEEYLVKQIDLSWGLMKNHVALACMIVIFLTITAYLKLFLKKYS 655  
Db 601 NPCNYATCTGEEYLVKQIDLSWGLMKNHVALACMIVIFLTITAYLKLFLKKYS 655

RESULT 4  
AA014781  
ID AA014781 standard; protein; 655 AA.  
XX  
AC AA014781;  
XX  
DT 28-JUN-2002 (first entry)  
XX  
DE Human BCRP protein.  
XX  
KW Human; BCRP protein; membrane penetrating region; cancer.  
XX  
OS Homo sapiens.  
XX  
PN JP2002065277-A.  
XX  
PD 05-MAR-2002.  
XX

241 SIHQPRYSIFKLFDSLTLASGRIMFHPGPAQEAALGYFESAGYHCEAYNNPADFFLDIING 300  
249 SIHQPRYSIFKLFDSLTLASGRIMFHPGPAQEAALGYFESAGYHCEAYNNPADFFLDIING 308  
301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIEIYVNSSFYKETKAEHLQSGGEKKK 360  
309 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIEIYVNSSFYKETKAEHLQSGGEKKK 368  
361 ITVFEKISYTTSFCHQLRWVSQRSFKNLLGNPQASIAQIITVTVGLVIGAIYFGLKXDS 420  
369 ITVFEKISYTTSFCHQLRWVSQRSFKNLLGNPQASIAQIITVTVGLVIGAIYFGLKXDS 428  
421 TGIQNRAGVLFFLTNNQCFSSVAVELFVVEKKLFHEIYISGYRVSSYFLGKLLSDLLP 480  
429 TGIQNRAGVLFFLTNNQCFSSVAVELFVVEKKLFHEIYISGYRVSSYFLGKLLSDLLP 488  
481 MTMLPSIIFTCIVVFMGLKPKADAFVVMFTLMVAYSASSMALAIAAGQSVSVATLL 540  
489 MTMLPSIIFTCIVVFMGLKPKADAFVVMFTLMVAYSASSMALAIAAGQSVSVATLL 548  
541 MTICFVFMIFPSGLLVNLTITIASWLSWLYQFSIPRYGFTALQHNBEFLGQFCPLNATGN 600  
549 MTICFVFMIFPSGLLVNLTITIASWLSWLYQFSIPRYGFTALQHNBEFLGQFCPLNATGN 608  
601 NPCNYATCTGEEYLVKQIDLSWGLMKNHVALACMIVIFLTITAYLKLFLKKYS 655  
609 NPCNYATCTGEEYLVKQIDLSWGLMKNHVALACMIVIFLTITAYLKLFLKKYS 663

RESULT 3  
AAB60104  
ID AAB60104 standard; protein; 655 AA.  
XX  
AC AAB60104;  
XX  
DT 28-MAR-2001 (first entry)  
XX  
DE Human transport protein TPPT-24.  
XX  
KW Human; transport protein; TPPT; transport disorder; metabolic disorder;  
KW neurological disorder; cardiovascular disorder; reproductive disorder;  
KW immune disorder; cancer.  
XX  
OS Homo sapiens.  
XX  
PN WO200078953-A2.  
XX  
PD 28-DEC-2000.  
XX  
PF 16-JUN-2000; 2000WO-US016668.  
XX  
PR 17-JUN-1999; 99US-0139923P.  
PR 10-AUG-1999; 99US-0148177P.  
PR 18-AUG-1999; 99US-0149357P.  
PR 28-OCT-1999; 99US-0162287P.  
XX  
(INCY-) INCYTE GENOMICS INC.  
XX  
Lal P, Yang J, Yue H, Hillman JL, Tang YT, Bandman O, Burford N;  
Baughn WF, Azimzai Y, Lu DAM, Au-Young J, Patterson C;  
WPI; 2001-041424/05.  
N-PSDB; AAF27724.  
XX  
Isolated polypeptide with a human transport protein sequence is useful  
PT for the diagnosis, prevention and treatment of disorders associated with  
PT the immune, reproductive and cardiovascular systems.  
XX  
Claim 2; Page 126-127; 165pp; English.  
XX  
The present invention provides the protein and coding sequences for 43  
CC novel human transport proteins (designated TPPTs). These can be used in



PF 31-AUG-2000; 2000JP-00263742.  
 XX 31-AUG-2000; 2000JP-00263742.  
 PR (GANK-) ZH GAN KENKYUKAI.  
 XX WPI; 2002-324198/36.  
 DR N-PSDB; AAL42412.  
 XX Mutant BCRP protein useful for treatment of cancer.  
 PT Claim 13; Page 7-8; 15pp; Japanese.  
 XX The invention comprises a mutant human BCRP protein, having a deletion,  
 CC replacement or addition of at least one amino acid in the fifth membrane  
 CC penetrating region of the wild-type BCRP protein. The mutant BCRP protein  
 CC can be used for the treatment of cancer. The present amino acid sequence  
 CC represents a human BCRP protein  
 XX Sequence 655 AA;  
 SQ

Query Match 99.8%; Score 3346; DB 5; Length 655;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 654; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MSSNVEVFPVSGQNTGFPATASNDLKAFTEGAVLSFHNICVRVKLSGFLPCRKPE 60  
 Db 1 MSSNVEVFPVSGQNTGFPATASNDLKAFTEGAVLSFHNICVRVKLSGFLPCRKPE 60  
 Qy 61 KEILSNINGIMKPGNALGPTGGKSLDLVLAARKDPSGLSDVLINGAPRANFKCN 120  
 Db 61 KEILSNINGIMKPGNALGPTGGKSLDLVLAARKDPSGLSDVLINGAPRANFKCN 120  
 Qy 121 SGYVQDDVVMGTLTVRENLFSAALRLATTMTNHEKNERINRVIOELGLDKVADSKVGT 180  
 Db 121 SGYVQDDVVMGTLTVRENLFSAALRLATTMTNHEKNERINRVIOELGLDKVADSKVGT 180  
 Qy 181 QFIRGVSGERKRTSIGMELTIDPSILFDEPTTGLDSSANAVLLKRMKSQGRITIF 240  
 Db 181 QFIRGVSGERKRTSIGMELTIDPSILFDEPTTGLDSSANAVLLKRMKSQGRITIF 240  
 Qy 241 SIHQPRYSIFKLFDSLTLLASGRLMFHGPAQALGFESAGYHCEAYNNPADPFLDIING 300  
 Db 241 SIHQPRYSIFKLFDSLTLLASGRLMFHGPAQALGFESAGYHCEAYNNPADPFLDIING 300  
 Qy 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIIYVNSFYKETAELHQLSGEKKKK 360  
 Db 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIIYVNSFYKETAELHQLSGEKKKK 360  
 Qy 361 ITVFKIEISYTSFCHQLRWKRSFKNLLGNPQASIAQIIVTVVLGLVIGAIYFGLKND 420  
 Db 361 ITVFKIEISYTSFCHQLRWKRSFKNLLGNPQASIAQIIVTVVLGLVIGAIYFGLKND 420  
 Qy 421 TGIQNRAGVLPFLTNQCFSSVAVELFVVEKKLFIHEYISGYRVSSYFLGKLLSLLP 480  
 Db 421 TGIQNRAGVLPFLTNQCFSSVAVELFVVEKKLFIHEYISGYRVSSYFLGKLLSLLP 480  
 Qy 481 MWMPLSIIFTCTIVYFMGLGPKADAFVMMFTLMWVAYSASSMALAIAGOSVVSVALTL 540  
 Db 481 MWMPLSIIFTCTIVYFMGLGPKADAFVMMFTLMWVAYSASSMALAIAGOSVVSVALTL 540  
 Qy 541 MTICFVFMWIFSGLLVNLTTIASWLSWLYQPSIPRYGFTALQHNFEFLQNCPCGLNATGN 600  
 Db 541 MTICFVFMWIFSGLLVNLTTIASWLSWLYQPSIPRYGFTALQHNFEFLQNCPCGLNATGN 600  
 Qy 601 NPCNATCTGEBYLVKQGDLSFGLWKNHVALACWIFVTIAYKLLFLKKYS 655  
 Db 601 NPCNATCTGEBYLVKQGDLSFGLWKNHVALACWIFVTIAYKLLFLKKYS 655

RESULT 5  
 AAU80028  
 ID AAU80028 standard; protein; 655 AA.

XX AAU80028;  
 XX 15-JUL-2002 (first entry)  
 XX Human ABCG2.  
 XX Human; ABCG2; transporter protein; anticancer drug tolerance;  
 XX indocarbazole.  
 XX Homo sapiens.  
 XX WO200228894-A1.  
 XX 11-APR-2002.  
 XX 18-SEP-2001; 2001WO-JP008112.  
 XX 03-OCT-2000; 2000JP-00303441.  
 XX (BANY ) BANYU PHARM CO LTD.  
 XX Komatani H, Hara Y, Kotani H, Nakagawa R;  
 XX WPI; 2002-352228/38.  
 XX N-PSDB; ABK49901.  
 XX ABCG2 gene encoding transporter protein capable of selectively  
 PT transporting indocarbazole compounds, useful in screening inhibitors and  
 PT anticancer agents for administration in chemotherapy.  
 XX Claim 1; Page 71-76; 98pp; Japanese.  
 XX The invention relates to an ABCG2 gene encoding a transporter protein  
 CC capable of imparting tolerance to an anticancer agent in mammals  
 CC comprising a fully defined sequence as given in the specification or an  
 CC amino acid sequence based on the sequence but with some amino acids  
 CC substituted, deleted or added. The gene and encoded protein are useful in  
 CC screening inhibitors and anticancer agents for administration in  
 CC chemotherapy with enhancement in sensitivity of cancer cell tolerance.  
 CC The gene relating to drug tolerance can be modified e.g. with the  
 CC transporter inhibitors, screened compounds, antibodies and antisense  
 CC nucleotides. The transporter is capable of selectively transporting  
 CC indocarbazole compounds extracellularly. The present sequence represents  
 CC the amino acid sequence of human ABCG2 protein  
 XX Sequence 655 AA;  
 SQ

Query Match 99.8%; Score 3346; DB 5; Length 655;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 654; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MSSNVEVFPVSGQNTGFPATASNDLKAFTEGAVLSFHNICVRVKLSGFLPCRKPE 60  
 Db 1 MSSNVEVFPVSGQNTGFPATASNDLKAFTEGAVLSFHNICVRVKLSGFLPCRKPE 60  
 Qy 61 KEILSNINGIMKPGNALGPTGGKSLDLVLAARKDPSGLSDVLINGAPRANFKCN 120  
 Db 61 KEILSNINGIMKPGNALGPTGGKSLDLVLAARKDPSGLSDVLINGAPRANFKCN 120  
 Qy 121 SGYVQDDVVMGTLTVRENLFSAALRLATTMTNHEKNERINRVIOELGLDKVADSKVGT 180  
 Db 121 SGYVQDDVVMGTLTVRENLFSAALRLATTMTNHEKNERINRVIOELGLDKVADSKVGT 180  
 Qy 181 QFIRGVSGERKRTSIGMELTIDPSILFDEPTTGLDSSANAVLLKRMKSQGRITIF 240  
 Db 181 QFIRGVSGERKRTSIGMELTIDPSILFDEPTTGLDSSANAVLLKRMKSQGRITIF 240  
 Qy 241 SIHQPRYSIFKLFDSLTLLASGRLMFHGPAQALGFESAGYHCEAYNNPADPFLDIING 300  
 Db 241 SIHQPRYSIFKLFDSLTLLASGRLMFHGPAQALGFESAGYHCEAYNNPADPFLDIING 300  
 Qy 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIIYVNSFYKETAELHQLSGEKKKK 360

Db 301 DSTAVLNREEDFKATEIIEPSKODKPLIEKLAIEYVNSFFYKETAELHQLSGGKXXX 360  
Qy 361 ITVFKESIVTTSFCHOLRWVSKSPKMLGNPOASIAQIIIVTVGLVIGALYFGLKND 420  
Db 361 ITVFKESIVTTSFCHOLRWVSKSPKMLGNPOASIAQIIIVTVGLVIGALYFGLKND 420  
Qy 421 TGIONRAGVLFLLTNNQCFSSVAVELFVVEKKLFTHIEYISGYRVSSYFLGKLSDLLP 480  
Db 421 TGIONRAGVLFLLTNNQCFSSVAVELFVVEKKLFTHIEYISGYRVSSYFLGKLSDLLP 480  
Qy 481 MTMLPSIIFTCIIVYFVLMGLKPKADAFVVMFTLMVAYSASSMALAIAAGQSVSVATLL 540  
Db 481 MTMLPSIIFTCIIVYFVLMGLKPKADAFVVMFTLMVAYSASSMALAIAAGQSVSVATLL 540  
Qy 541 MTICFVFMIFSGLLVNLTTIASWLSWLYFSIPRYGFTALQHNELGONFCPLNATGN 600  
Db 541 MTICFVFMIFSGLLVNLTTIASWLSWLYFSIPRYGFTALQHNELGONFCPLNATGN 600  
Qy 601 NPCNYATCTGEEYLVKQIDLSPWGLWKNHVALACMIVIFLTAYKLLFLKKYS 655  
Db 601 NPCNYATCTGEEYLVKQIDLSPWGLWKNHVALACMIVIFLTAYKLLFLKKYS 655

RESULT 6  
ABR58077  
ID ABR58077 standard; protein; 655 AA.  
XX  
AC ABR58077;  
XX  
DT 15-OCT-2003 (first entry)  
XX  
DE Human ABCG2 protein.  
XX  
KW ABCG2; antidiabetic; cell therapy; diabetes mellitus;  
KW pancreatic stem cell; islets of langerhans; insulin.  
XX  
OS Homo sapiens.  
XX  
PN WO2003026584-A2.  
XX  
PD 03-APR-2003.  
XX  
XX 26-SEP-2002; 2002WO-US030700.  
XX  
XX 26-SEP-2001; 2001US-00963875.  
XX  
XX 11-APR-2002; 2002US-00120687.  
XX  
XX 02-MAY-2002; 2002US-00136891.  
XX  
XX (GEO ) GEN HOSPITAL CORP.  
XX  
XX Habener JF, Zulewski H, Thomas MK, Abraham EJ, Vallejo M;  
XX  
XX Leesch CA, Nolan AL, Lechner A;  
XX  
XX WPI; 2003-354625/33.  
XX  
XX N-PSDB; ACC80605.

PT Treating a patient with diabetes mellitus by isolating a nestin- or ABCG2  
PT -positive pancreatic stem cell from a pancreatic islet of a donor and  
PT transferring the stem cell into the patient.  
XX  
XX Disclosure; Fig 18B; 107pp; English.  
XX  
XX The invention relates to a method of treating a patient with diabetes  
XX mellitus by isolating a nestin- or ABCG2-positive pancreatic stem cell  
XX from a pancreatic islet of a donor, and transferring the stem cell into  
XX the patient whereby the stem cell differentiates into an insulin-  
XX producing cell. Alternatively, the nestin- or ABCG2-positive stem is  
XX induced into a pancreatic progenitor cell prior to isolation and  
XX transfer. This sequence corresponds to the human ABCG2 protein and the  
XX encoding gene is detected in the method of the invention. The method is  
XX useful for preparing a pharmaceutical composition for treating diabetes  
XX mellitus. The stem cells can be further characterised for correct gene

CC expression using the primers and probes ACC80607-ACC80671  
XX Sequence 655 AA;

Query Match 99.8%; Score 3346; DB 6; Length 655;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 654; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MSSNVVEVFPVSGQNTNGFPATASNDLKAFTGAVLSFNHCYRVKLSGFLPCRKPE 60  
Db 1 MSSNVVEVFPVSGQNTNGFPATASNDLKAFTGAVLSFNHCYRVKLSGFLPCRKPE 60  
Qy 61 KEILSNINGIMKPGNALIGPTGGKSSLLDLAARKDPSGLSDVNLINGAPRANFKCN 120  
Db 61 KEILSNINGIMKPGNALIGPTGGKSSLLDLAARKDPSGLSDVNLINGAPRANFKCN 120  
Qy 121 SGYVQDDVVGTLTVRENIQFSAALRLATMTNHEKNERINRVIQELGDKVADSKVGT 180  
Db 121 SGYVQDDVVGTLTVRENIQFSAALRLATMTNHEKNERINRVIQELGDKVADSKVGT 180  
Qy 181 QFIRGSGGGRKRTSIGMELITDPSILFLDBPTTGLDSSSTANAVLILLKMSKQRTIIF 240  
Db 181 QFIRGSGGGRKRTSIGMELITDPSILFLDBPTTGLDSSSTANAVLILLKMSKQRTIIF 240  
Qy 241 SIHQPRYSIFKLFDSLTLASGRMLMFGPAQALGYFESAGYHCEAYNNPADFFLDIING 300  
Db 241 SIHQPRYSIFKLFDSLTLASGRMLMFGPAQALGYFESAGYHCEAYNNPADFFLDIING 300  
Qy 301 DSTAVLNREEDFKATEIIEPSKODKPLIEKLAIEYVNSFFYKETAELHQLSGGKXXX 360  
Db 301 DSTAVLNREEDFKATEIIEPSKODKPLIEKLAIEYVNSFFYKETAELHQLSGGKXXX 360  
Qy 361 ITVFKESIVTTSFCHOLRWVSKSPKMLGNPOASIAQIIIVTVGLVIGALYFGLKND 420  
Db 361 ITVFKESIVTTSFCHOLRWVSKSPKMLGNPOASIAQIIIVTVGLVIGALYFGLKND 420  
Qy 421 TGIONRAGVLFLLTNNQCFSSVAVELFVVEKKLFTHIEYISGYRVSSYFLGKLSDLLP 480  
Db 421 TGIONRAGVLFLLTNNQCFSSVAVELFVVEKKLFTHIEYISGYRVSSYFLGKLSDLLP 480  
Qy 481 MTMLPSIIFTCIIVYFVLMGLKPKADAFVVMFTLMVAYSASSMALAIAAGQSVSVATLL 540  
Db 481 MTMLPSIIFTCIIVYFVLMGLKPKADAFVVMFTLMVAYSASSMALAIAAGQSVSVATLL 540  
Qy 541 MTICFVFMIFSGLLVNLTTIASWLSWLYFSIPRYGFTALQHNELGONFCPLNATGN 600  
Db 541 MTICFVFMIFSGLLVNLTTIASWLSWLYFSIPRYGFTALQHNELGONFCPLNATGN 600  
Qy 601 NPCNYATCTGEEYLVKQIDLSPWGLWKNHVALACMIVIFLTAYKLLFLKKYS 655  
Db 601 NPCNYATCTGEEYLVKQIDLSPWGLWKNHVALACMIVIFLTAYKLLFLKKYS 655

RESULT 7  
ADA10917  
ID ADA10917 standard; protein; 655 AA.  
XX  
AC ADA10917;  
XX

DT 06-NOV-2003 (first entry)  
XX  
XX Human cDNA differentially expressed in colon cancer #23 product.  
XX  
XX differential expression; colon cancer; cancer; human.  
XX  
XX Homo sapiens.  
XX  
XX US2002160382-A1.  
XX  
XX 31-OCT-2002.  
XX  
XX 11-OCT-2001; 2001US-00981353.  
XX

PR 11-OCT-2000; 2000US-0239841P.  
XX (LASE/) LASEK A W.  
PA (JONE/) JONES D A.  
XX Lasek AW, Jones DA;  
XX WPI; 2003-265756/26.  
DR N-PSDB; ADA10916.  
XX New combination comprising cDNAs that are differentially expressed in  
PT colon disorder, useful for diagnosing, treating, staging or monitoring  
PT treatment for colon cancer.  
XX Example 14; SEQ ID NO 35; 231pp; English.  
XX The invention relates to a combination comprising cDNAs that are  
CC differentially expressed in colon disorder. The methods and compositions  
CC of the present invention are useful for diagnosing, treating, staging or  
CC monitoring treatment for colon cancer. They are also useful in high  
CC throughput methods for using cDNAs to detect differential expression of  
CC nucleic acids in a sample, screening molecules or compounds to identify a  
CC ligand which specifically binds a cDNA and using a protein to screen  
CC molecules or compounds to identify at least one ligand which specifically  
CC binds the protein. The present sequence represents the amino acid  
CC sequence of a human cDNA differentially expressed in colon cancer  
CC protein.  
XX SQ Sequence 655 AA;  
Query Match 99.8%; Score 3346; DB 6; Length 655;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 654; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MSSNVEVPIPVSOQNTNGFPATASNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPYE 60  
Db 1 MSSNVEVPIPVSOQNTNGFPATASNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPYE 60  
QY 61 KEILSNINGIMKPGNALILGPTGGKSSLLDVLARKDPSGLSDVINGAPRANFKCN 120  
Db 61 KEILSNINGIMKPGNALILGPTGGKSSLLDVLARKDPSGLSDVINGAPRANFKCN 120  
QY 121 SGYVQDDVVMGTLTVRENLOPSAALRLATWTNHEKNERINRVIOELGLDKVADSKVGT 180  
Db 121 SGYVQDDVVMGTLTVRENLOPSAALRLATWTNHEKNERINRVIOELGLDKVADSKVGT 180  
QY 181 QIRGVSGGERKRTSIGMELITDPSILFLDPTTGLDSTANAVLLLLKRMKSGKRTIIF 240  
Db 181 QIRGVSGGERKRTSIGMELITDPSILFLDPTTGLDSTANAVLLLLKRMKSGKRTIIF 240  
QY 241 SIHQPRYSIFKFLDLSLTLASGRMLPHGPAQALGYFESAGYHCEAYNNPADFLDIING 300  
Db 241 SIHQPRYSIFKFLDLSLTLASGRMLPHGPAQALGYFESAGYHCEAYNNPADFLDIING 300  
QY 301 DSTAVALNEEDFKATEIIEPSKQDKPLEKLAIEYVNSFFYKETAELHOLSGEKKKK 360  
Db 301 DSTAVALNEEDFKATEIIEPSKQDKPLEKLAIEYVNSFFYKETAELHOLSGEKKKK 360  
QY 361 ITVFKESYTTGFCQHLRWVSKRSPKLLGNPQASIAQIIVTVLGLVIGAIYFGLKNDS 420  
Db 361 ITVFKESYTTGFCQHLRWVSKRSPKLLGNPQASIAQIIVTVLGLVIGAIYFGLKNDS 420  
QY 421 TGIONPAGVLFLTNNOCSSYSAVELFVVEKKLFTHIYISGYVSVSYFGLKLSLLP 480  
Db 421 TGIONPAGVLFLTNNOCSSYSAVELFVVEKKLFTHIYISGYVSVSYFGLKLSLLP 480  
QY 481 MTPLPSIIFTCIVYFMGLKPKADAFFVMMFTIMMVAYSASSMALAIAAGQSVSVATLL 540  
Db 481 MRMLPSIIFTCIVYFMGLKPKADAFFVMMFTIMMVAYSASSMALAIAAGQSVSVATLL 540  
QY 541 MTICFVMMIFSGLLVNLITIASLSWLQYFIPRYGFTALOHNEFLQNCPCPLNATGN 600  
Db 541 MTICFVMMIFSGLLVNLITIASLSWLQYFIPRYGFTALOHNEFLQNCPCPLNATGN 600

QY 601 NPCNYATCTGBEYLVKQIGIDLSPGMLWKNHVALACMIVIFLTIAVLKLLFLKKYS 655  
Db 601 NPCNYATCTGBEYLVKQIGIDLSPGMLWKNHVALACMIVIFLTIAVLKLLFLKKYS 655  
RESULT 8  
ADC54182  
ID ADC54182 standard; protein; 655 AA.  
XX AC ADC54182;  
XX DT 18-DEC-2003 (first entry)  
XX DE Human breast cancer resistance protein (BCRP) amino acid sequence.  
XX KW cancer cell; anti-cancer agent; steroid hormone; oestrogenic effect;  
KW BCRP; breast cancer resistance protein; cytostatic; camptothecins;  
KW mitoxantrone; 7-hydroxy staurosporine; adriamycin; cancer chemotherapy;  
KW human.  
XX OS Homo sapiens.  
XX PN JP2003063989-A.  
XX PD 05-MAR-2003.  
XX PF 23-AUG-2001; 2001JP-00252953.  
XX PR 23-AUG-2001; 2001JP-00252953.  
XX PA (GANK-) ZH GAN KENYUKAI.  
XX DR WPI; 2003-735321/70.  
XX N-PSDB; ADC54181.  
PT Agent that overcomes resistance of cancer cell against anti-cancer agent,  
PT comprises a steroid hormone, or a compound which exhibits antagonistic  
PT activity against the hormone, with the cancer cell expressing BCRP gene.  
XX Example 1; SEQ ID NO 4; 15pp; Japanese.  
XX This invention relates to a novel agent which overcomes resistance of a  
CC cancer cell against an anti-cancer agent (AA), comprising as an active  
CC ingredient a steroid hormone, a compound having oestrogenic effect, or a  
CC compound which exhibits antagonistic activity against the hormone, where  
CC the cancer cell expresses the BCRP (breast cancer resistance protein)  
CC gene. The agent of the invention may have cytostatic activity. The  
CC invention is useful for overcoming resistance of a cancer against an anti  
CC -cancer agent such as camptothecin, mitoxantrone, 7-hydroxy  
CC staurosporine and adriamycin. The therapeutic effective anti-cancer agent  
CC is recovered, due to the use of the agent of the invention. Also the  
CC dosages of anti-cancer agent can be maintained easily, and adverse  
CC effects of cancer chemotherapy can be suppressed. The present sequence is  
CC that of the human BCRP protein which was used to develop the novel agent  
CC of the invention.  
XX SQ Sequence 655 AA;  
Query Match 99.8%; Score 3346; DB 7; Length 655;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 654; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MSSNVEVPIPVSOQNTNGFPATASNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPYE 60  
Db 1 MSSNVEVPIPVSOQNTNGFPATASNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPYE 60  
QY 61 KEILSNINGIMKPGNALILGPTGGKSSLLDVLARKDPSGLSDVINGAPRANFKCN 120  
Db 61 KEILSNINGIMKPGNALILGPTGGKSSLLDVLARKDPSGLSDVINGAPRANFKCN 120  
QY 121 SGYVQDDVVMGTLTVRENLOPSAALRLATWTNHEKNERINRVIOELGLDKVADSKVGT 180  
Db 121 SGYVQDDVVMGTLTVRENLOPSAALRLATWTNHEKNERINRVIOELGLDKVADSKVGT 180

Db 121 SGVVQDDVVMGTLTVRENLFQSAALRLATTWTHNEKNERINRVIOELGLDKVADSKVGT 180  
 QY 181 QFIRGVSGGERKRTSIGMELITDPSILFDLPTTGLDSDSTANAVLLLLKRMKSKQGRITIF 240  
 Db 181 QFIRGVSGGERKRTSIGMELITDPSILFDLPTTGLDSDSTANAVLLLLKRMKSKQGRITIF 240  
 QY 241 SIHQPRYSIFKLFDSLTLLASGRLMFHGPAQALGYFESAGVHCEAYNNPADFFLDIING 300  
 Db 241 SIHQPRYSIFKLFDSLTLLASGRLMFHGPAQALGYFESAGVHCEAYNNPADFFLDIING 300  
 QY 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAETIYVNSSFYKETAELHQLSGEKKKK 360  
 Db 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAETIYVNSSFYKETAELHQLSGEKKKK 360  
 QY 361 ITVFKETSYTTSFCHOLRWVSKRSFKNLLGNPOASIAQIIVTVLGLVIGAIYFGLKND 420  
 Db 361 ITVFKETSYTTSFCHOLRWVSKRSFKNLLGNPOASIAQIIVTVLGLVIGAIYFGLKND 420  
 QY 421 TGIQNRAGVLFLLTNQCFSSVSAVELFVVEKKLFIHEYISGYRVSSYFGLKLLSDLLP 480  
 Db 421 TGIQNRAGVLFLLTNQCFSSVSAVELFVVEKKLFIHEYISGYRVSSYFGLKLLSDLLP 480  
 QY 481 MTWLPSSIIFTCIVYFVLMGLKPKADAFVMMFTLMVAYSSASSMALAIAGQSVSVATLL 540  
 Db 481 MRMLPSSIIFTCIVYFVLMGLKPKADAFVMMFTLMVAYSSASSMALAIAGQSVSVATLL 540  
 QY 541 MTICFVFMIFSGLLVNLTTIASWLSWLYQFSIPRYGFTALQHNFLQGNFCPLNATGN 600  
 Db 541 MTICFVFMIFSGLLVNLTTIASWLSWLYQFSIPRYGFTALQHNFLQGNFCPLNATGN 600  
 QY 601 NPCNYATCTGEEYLVKQIDLSFWGLWKNHVALACMIVIFLTAYLKLFLKKYS 655  
 Db 601 NPCNYATCTGEEYLVKQIDLSFWGLWKNHVALACMIVIFLTAYLKLFLKKYS 655

## RESULT 9

AAO14783  
 ID AAO14783 standard; protein; 665 AA.

XX AAO14783;

AC 28-JUN-2002 (first entry)

DT Human BCRP-related protein 2.

DE Human; BCRP-related protein; membrane penetrating region; cancer.

XX Homo sapiens.

XX JP2002065277-A.

XX 05-MAR-2002.

XX 31-AUG-2000; 2000JP-00263742.

XX 31-AUG-2000; 2000JP-00263742.

XX (GANK-) ZH GAN KENYUKAI.

XX WPI; 2002-324198/36.

DR N-PSDB; AAL42414.

XX Mutant BCRP protein useful for treatment of cancer.

XX Example 3; Page 12-13; 15pp; Japanese.

XX The invention comprises a mutant human BCRP protein, having a deletion, replacement or addition of at least one amino acid in the fifth membrane penetrating region of the wild-type BCRP protein. The mutant BCRP protein can be used for the treatment of cancer. The present amino acid sequence represents a human BCRP-related protein

XX Sequence 665 AA;

PR 31-MAY-2000; 2000US-00584586.

Query Match 99.6%; Score 3339; DB 5; Length 665;  
 Best Local Similarity 99.7%; Pred. No. 0;  
 Matches 653; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSSSNVEFIPVSGQNTNGFPATASNDLKAFTEGAVLSFPHNICYRVKLSGFLPCRKPVE 60  
 Db 11 MSSSNVEFIPVSGQNTNGFPATASNDLKAFTEGAVLSFPHNICYRVKLSGFLPCRKPVE 70  
 QY 61 KEILSNINGIMKPLGNAIILGPTGGKSSLLDLVLAARKDPSGLSGDVLINGAPRPANFKCN 120  
 Db 71 KEILSNINGIMKPLGNAIILGPTGGKSSLLDLVLAARKDPSGLSGDVLINGAPRPANFKCN 130  
 QY 121 SGVVQDDVVMGTLTVRENLFQSAALRLATTWTHNEKNERINRVIOELGLDKVADSKVGT 180  
 Db 131 SGVVQDDVVMGTLTVRENLFQSAALRLATTWTHNEKNERINRVIOELGLDKVADSKVGT 190  
 QY 181 QFIRGVSGGERKRTSIGMELITDPSILFDLPTTGLDSDSTANAVLLLLKRMKSKQGRITIF 240  
 Db 191 QFIRGVSGGERKRTSIGMELITDPSILFDLPTTGLDSDSTANAVLLLLKRMKSKQGRITIF 250  
 QY 241 SIHQPRYSIFKLFDSLTLLASGRLMFHGPAQALGYFESAGVHCEAYNNPADFFLDIING 300  
 Db 251 SIHQPRYSIFKLFDSLTLLASGRLMFHGPAQALGYFESAGVHCEAYNNPADFFLDIING 310  
 QY 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAETIYVNSSFYKETAELHQLSGEKKKK 360  
 Db 311 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAETIYVNSSFYKETAELHQLSGEKKKK 370  
 QY 361 ITVFKETSYTTSFCHOLRWVSKRSFKNLLGNPOASIAQIIVTVLGLVIGAIYFGLKND 420  
 Db 371 ITVFKETSYTTSFCHOLRWVSKRSFKNLLGNPOASIAQIIVTVLGLVIGAIYFGLKND 430  
 QY 421 TGIQNRAGVLFLLTNQCFSSVSAVELFVVEKKLFIHEYISGYRVSSYFGLKLLSDLLP 480  
 Db 431 TGIQNRAGVLFLLTNQCFSSVSAVELFVVEKKLFIHEYISGYRVSSYFGLKLLSDLLP 490  
 QY 481 MTWLPSSIIFTCIVYFVLMGLKPKADAFVMMFTLMVAYSSASSMALAIAGQSVSVATLL 540  
 Db 491 MRMLPSSIIFTCIVYFVLMGLKPKADAFVMMFTLMVAYSSASSMALAIAGQSVSVATLL 550  
 QY 541 MTICFVFMIFSGLLVNLTTIASWLSWLYQFSIPRYGFTALQHNFLQGNFCPLNATGN 600  
 Db 551 MTICFVFMIFSGLLVNLTTIASWLSWLYQFSIPRYGFTALQHNFLQGNFCPLNATGN 610  
 QY 601 NPCNYATCTGEEYLVKQIDLSFWGLWKNHVALACMIVIFLTAYLKLFLKKYS 655  
 Db 611 NPCNYATCTGEEYLVKQIDLSFWGLWKNHVALACMIVIFLTAYLKLFLKKYS 665

## RESULT 10

ABB07273  
 ID ABB07273 standard; protein; 655 AA.

XX ABB07273;

XX 26-MAR-2002 (first entry)

XX Human BCRP (huBCRP) sequence.

XX Stem cell; ATP transport protein; ATP-binding cassette; antiparkinsonian;

KW hepatotropic; neurodegenerative; cytosolic; antianemic; muscular; BCRP;

KW cardiant; gene therapy.

XX Homo sapiens.

XX WO200192877-A2.

XX 06-DEC-2001.

XX 30-MAY-2001; 2001WO-US017459.

XX 31-MAY-2000; 2000US-00584586.

29-MAY-2001; 2001US-00866866.  
(STUD-) ST JUDE CHILDREN'S RES HOSPITAL.  
Sorrentino B, Schuetz J;  
WPI; 2002-114368/15.  
N-PSDB; ABA94383.  
Identifying a stem cell, for treating e.g., muscular dystrophy,  
myocardial infarction, Parkinson's disease, or neurodegenerative  
disorders, comprises detecting the expression of an ATP transport protein  
(BCRP) by a cell.  
Example 1; Page 86-87; 87pp; English.  
The invention provides a method of identifying and/or isolating a stem  
cell that involves detecting the expression of an ATP transport protein  
containing a conserved ATP-binding cassette (ABCP) by a cell in a sample  
comprising stem cells. The isolated stem cells may be used in the  
treatment of diseases such as muscular dystrophy, degenerative liver  
disorder, myocardial infarction, Parkinson's disease, degenerative  
disorders of the brain, and for tissue regeneration or replacement.  
Haematopoietic cells can be used in bone marrow transplants (e.g., for  
treatment of leukemia) and for ex vivo gene therapy for treating blood  
diseases such as sickle cell anemia and thalassemia. The stem cells can  
also be used as cell targets in gene therapy protocols. The present  
sequence represents the human BCRP  
Sequence 655 AA;  
Query Match . 99.6%; Score 3338; DB 5; Length 655;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 653; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 MSSNVEVFPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICVRVKLSGFLPCRPVE 60  
Db 1 MSSNVEVFPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICVRVKLSGFLPCRPVE 60  
QY 61 KEILSNINGIMKPGNALIGPTGGKSSLLDVLAAKDPGSLGSDVLINGAPRANFKCN 120  
Db 61 KEILSNINGIMKPGNALIGPTGGKSSLLDVLAAKDPGSLGSDVLINGAPRANFKCN 120  
QY 121 SGYVQDDVVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIOELGLDKVADSKVGT 180  
Db 121 SGYVQDDVVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIOELGLDKVADSKVGT 180  
QY 181 QIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSDSTANAVLLILKMSKQGRITIF 240  
Db 181 QIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSDSTANAVLLILKMSKQGRITIF 240  
QY 241 SIHQPRYSIFKLFDSLTLLASGRLMFHGPAQALGYFESAGVHCEAYNNPADFFDIING 300  
Db 241 SIHQPRYSIFKLFDSLTLLASGRLMFHGPAQALGYFESAGVHCEAYNNPADFFDIING 300  
QY 301 DSTAVALNEEDFKATEIIEPSKQDKPLEKLAETVNSSFKETKAEHLHOLSGEKKK 360  
Db 301 DSTAVALNEEDFKATEIIEPSKQDKPLEKLAETVNSSFKETKAEHLHOLSGEKKK 360  
QY 361 ITVFEKISYTTSFCHQLRWVSKRSPKNLGNPQASIAQIIVTVLGLVIGAIYFGLKND 420  
Db 361 ITVFEKISYTTSFCHQLRWVSKRSPKNLGNPQASIAQIIVTVLGLVIGAIYFGLKND 420  
QY 421 TGIONRAGVLFLLTNQCESSVAVELFVVEKKLFIEHISGYRVSSVYFGLKLSLLP 480  
Db 421 TGIONRAGVLFLLTNQCESSVAVELFVVEKKLFIEHISGYRVSSVYFGLKLSLLP 480  
QY 481 MTMLPISIIITCIIVYFMGLKPKADAFVVMFTIMVAVSASSMALAIAAGQSVSVATLL 540  
Db 481 MTMLPISIIITCIIVYFMGLKPKADAFVVMFTIMVAVSASSMALAIAAGQSVSVATLL 540  
QY 541 MTICFVFMIFSGLLVNLTTIASWLSWLOQYFSPRYGFTALQHNELGQNCPLNATGN 600

Db 541 MTICFVFMIFSGLLVNLTTIASWLSWLOQYFSPRYGFTALQHNELGQNCPLNATGN 600  
QY 601 NPCNYATCTGEEYLVKQIDLSPWGLWKNHVALACMIVIFLTITAYLKLFLKKYS 655  
Db 601 NPCNYATCTGEEYLVKQIDLSPWGLWKNHVALACMIVIFLTITAYLKLFLKKYS 655

RESULT 11  
AA95365  
ID AA95365 standard; protein; 655 AA.

XX AC AA95365;  
XX DT 25-SEP-2000 (first entry)  
XX DE ATP-binding cassette protein MXR1.  
XX KW ATP-binding cassette protein; ABC protein; MXR1; human;  
XX KW cytoxin resistance; colon carcinoma; therapy.

XX OS Homo sapiens.  
XX PN WO200036101-A2.  
XX PD 22-JUN-2000.  
XX PF 24-NOV-1999; 99WO-US028107.  
XX PR 30-NOV-1998; 98US-0110473P.  
XX PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX PI Dean M, Allikmets R, Bates SE, Fojo AT;  
XX DR WPI; 2000-442385/38.  
XX DR N-PSDB; AAR27938.

XX PT Adenosine triphosphate (ATP)-binding cassette proteins and their nucleic  
XX PT acids, useful for identifying agents that may be used to treat colonic  
XX PS carcinomas resistant to cytotoxic therapeutic agents.  
XX PS Claim 1(ii); Page 46; 49pp; English.

XX CC The present sequence is that of an ATP-binding cassette protein (ABC  
XX CC protein) designated MXR1 that confers mitoxantrone resistance to SI-M1-80  
XX CC human colon carcinoma cells. Nucleic acids encoding the ABC protein may  
XX CC be used to produce the protein according to standard recombinant DNA  
XX CC methodologies. The expressed proteins may be used to confer resistance to  
XX CC cytotoxins such as mitoxantrone and/or daunomycin to the host cells in  
XX CC which they are expressed. In disease conditions such as carcinoma of the  
XX CC colon, breast and gastrointestinal tract, the ABC protein may be  
XX CC expressed to provide resistance to cytotoxic therapeutic agents. The  
XX CC nucleic acids and proteins may be used to produce host cell models of  
XX CC resistant cells which can be used to screen for candidate agents that  
XX CC inhibit the expression and/or activity of the ABC protein. These agents  
XX CC may be used to down regulate ABC protein expression in carcinoma cells  
XX CC and sensitize them to cytotoxic therapeutic agents

XX SQ Sequence 655 AA;

Query Match 99.4%; Score 3331; DB 3; Length 655;  
Best Local Similarity 99.4%; Pred. No. 0;  
Matches 651; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
QY 1 MSSNVEVFPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICVRVKLSGFLPCRPVE 60  
Db 1 MSSNVEVFPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICVRVKLSGFLPCRPVE 60  
QY 61 KEILSNINGIMKPGNALIGPTGGKSSLLDVLAAKDPGSLGSDVLINGAPRANFKCN 120  
Db 61 KEILSNINGIMKPGNALIGPTGGKSSLLDVLAAKDPGSLGSDVLINGAPRANFKCN 120  
QY 121 SGYVQDDVVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIOELGLDKVADSKVGT 180

```
Db 121 SGYVQDDVVMGTLTVRENLOFSAALRLATMTNHEKNERINRVIEELGLDKVADSKVGT 180
QY 181 QPIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVALLKRMKSKQGRITIF 240
Db 181 QPIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVALLKRMKSKQGRITIF 240
QY 241 SIHQPRYSIFKLFDSLTLLASGRMLMFGPAQALGYFESAGYHCEAYNNPADFFLDIING 300
Db 241 SIHQPRYSIFKLFDSLTLLASGRMLMFGPAQALGYFESAGYHCEAYNNPADFFLDIING 300
QY 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIIYVNSFFYKETKAEHLHQLSGGEKKK 360
Db 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIIYVNSFFYKETKAEHLHQLSGGEKKK 360
QY 361 ITVFKESIYTTSFCHQLRWWSKRSFKNLLGNPQASIAQIIVTVVLGLVIGAIYFGLKND 420
Db 361 ITVFKESIYTTSFCHQLRWWSKRSFKNLLGNPQASIAQIIVTVVLGLVIGAIYFGLKND 420
QY 421 TGQONRAGVLFLLTNNQCFSSYSAVELFVVEKKLFHEIYSGYRVSSYFLGKLSDDL 480
Db 421 TGQONRAGVLFLLTNNQCFSSYSAVELFVVEKKLFHEIYSGYRVSSYFLGKLSDDL 480
QY 481 MTMLPSIIFTCIYVFMGLKPKADAFVMMFTLMVAYSSASSMALAIAAGQSVSVATLL 540
Db 481 MRMLPSIIFTCIYVFMGLKPKADAFVMMFTLMVAYSSASSMALAIAAGQSVSVATLL 540
QY 541 MTICFVFMFTPSGLVNLATIASWLSWLOVESIPRYGTALOHNEFLGONFCPLNATGN 600
Db 541 MTICFVFMFTPSGLVNLATIASWLSWLOVESIPRYGTALOHNEFLGONFCPLNATGN 600
QY 601 NPCNYATCTGEEYLVKQGDLSWGLWQHVALACMIVIFLTAYLKLFLKKYS 655
Db 601 NPCNYATCTGEEYLVKQGDLSWGLWQHVALACMIVIFLTAYLKLFLKKYS 655

RESULT 12
AAU04348
ID AAU04348 standard; protein; 655 AA.
AC AAU04348;
XX
DT 23-OCT-2001 (first entry)
XX
DE Human BCRP/MXR/ABCP protein.
XX
KW ATP-binding cassette transporter protein-inhibiting peptide; ABC;
KW transmembrane domain; cancer; tumour; HIV; AIDS; BCRP/MXR/ABCP;
KW acquired immunodeficiency syndrome; human immunodeficiency syndrome.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Domain 398..420
FT FT /label= Transmembrane_domain_1
FT Domain 427..450
FT FT /label= Transmembrane_domain_2
FT Domain 478..501
FT FT /label= Transmembrane_domain_3
FT Domain 504..527
FT FT /label= Transmembrane_domain_4
FT Domain 539..558
FT FT /label= Transmembrane_domain_5
FT Domain 629..650
FT FT /label= Transmembrane_domain_6
XX
XX WO200136477-A2.
XX
XX 25-MAY-2001.
XX
XX 17-NOV-2000; 2000WO-US031817.
XX
XX 13-NOV-1999; 99US-0166382P.
PR
```

```
PR 22-NOV-1999; 99US-0166767P.
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX Tarasova NI, Michejda CU, Gottesman MM, Hrycyna C;
XX WPI; 2001-381224/40.
XX New ATP-binding cassette transporter-inhibiting peptides derived from
XX transmembrane domain of the transporter, useful for inhibiting biological
XX activity of the transporter and sensitizing cancer cells to
XX chemotherapeutics.
XX Disclosure; Fig 2; 89pp; English.
XX The sequence represents human BCRP/MXR/ABCP protein, an ATP-binding
XX cassette (ABC) transporter protein. Peptides derived from its
XX transmembrane domains are used as ATP-binding cassette transporter
XX protein-inhibiting peptides. The peptides have a first end and a second
XX end, where at the first end there is a group that is negatively charged
XX under physiological conditions and at the second end a group that is
XX neutrally charged under physiological conditions and is at least 70%
XX identical to a transmembrane domain of an ABC transporter. The peptides
XX are used for inhibiting the biological activity, in particular ion flux
XX or translocation, cytochrome efflux or translocation, phosphorylation,
XX protein synthesis or degradation, cellular morphology, secretion,
XX production of particular components such as soluble inositol phosphates,
XX tumour growth, chemotaxis, mitogenic response, cell growth activation or
XX secretion of a target ATP-binding cassette (ABC) transporter. The
XX peptides are useful to sensitize cancer cells to standard
XX chemotherapeutics, which chemotherapeutics are then administered to kill
XX the cancer cells. Transmembrane analogues of ABC transporter protein are
XX administered to patients infected with HIV-1, who are taking a protease
XX inhibitor. The ability of HIV-1 infected cells to efflux the protease
XX inhibitor is impaired, resulting in higher concentration of the protease
XX inhibitor in the affected cell. This reduces the amount of protease
XX inhibitor administered to the patient to achieve a virus-inhibiting dose
XX in the infected cells
XX
XX SQ Sequence 655 AA;
XX
XX Query Match 99.4%; Score 3331; DB 4; Length 655;
XX Best Local Similarity 99.4%; Pred. No. 0;
XX Matches 651; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 MSSNVVEFIPVSGQNTGPFATASNDLKAPTEGAVLSFHNICVRVKLKGFLPCRPVE 60
Db 1 MSSNVVEFIPVSGQNTGPFATVSNLDLKAPTEGAVLSFHNICVRVKLKGFLPCRPVE 60
QY 61 KEILSNINGIMKPGNALIGPTGGKSSLLDVLAAKDPGLSGDLVINGAPRANFKCN 120
Db 61 KEILSNINGIMKPGNALIGPTGGKSSLLDVLAAKDPGLSGDLVINGAPRANFKCN 120
QY 121 SGYVQDDVVMGTLTVRENLOFSAALRLATMTNHEKNERINRVIEELGLDKVADSKVGT 180
Db 121 SGYVQDDVVMGTLTVRENLOFSAALRLATMTNHEKNERINRVIEELGLDKVADSKVGT 180
QY 181 QPIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVALLKRMKSKQGRITIF 240
Db 181 QPIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVALLKRMKSKQGRITIF 240
QY 241 SIHQPRYSIFKLFDSLTLLASGRMLMFGPAQALGYFESAGYHCEAYNNPADFFLDIING 300
Db 241 SIHQPRYSIFKLFDSLTLLASGRMLMFGPAQALGYFESAGYHCEAYNNPADFFLDIING 300
QY 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIIYVNSFFYKETKAEHLHQLSGGEKKK 360
Db 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIIYVNSFFYKETKAEHLHQLSGGEKKK 360
QY 361 ITVFKESIYTTSFCHQLRWWSKRSFKNLLGNPQASIAQIIVTVVLGLVIGAIYFGLKND 420
Db 361 ITVFKESIYTTSFCHQLRWWSKRSFKNLLGNPQASIAQIIVTVVLGLVIGAIYFGLKND 420
```

QY 421 TGIQNRAGVLPFLTTNOCFSSVSAVELFVVEKKLFIHEYISGYRVSSYFLGKLLSDLLP 480  
 DB 421 TGIQNRAGVLPFLTTNOCFSSVSAVELFVVEKKLFIHEYISGYRVSSYFLGKLLSDLLP 480  
 QY 481 MTMLPSIIFTCTIVFMGLGKPKADAFFVMMFTLMMVAYSASSMALAIAAGQSVSVATLL 540  
 DB 481 MTMLPSIIFTCTIVFMGLGKPKADAFFVMMFTLMMVAYSASSMALAIAAGQSVSVATLL 540  
 QY 541 MTICFVFMIFSGLLVNLTTIASLSWLQYFSIPRYGFTALQHNFFLQNFQCFGLNATGN 600  
 DB 541 MTICFVFMIFSGLLVNLTTIASLSWLQYFSIPRYGFTALQHNFFLQNFQCFGLNATGN 600  
 QY 601 NPCNYATCTGSEYLVKQIDLSPLGMLKQNHVALACMIVIFLTIAYLKLLFLKKYS 655  
 DB 601 NPCNYATCTGSEYLVKQIDLSPLGMLKQNHVALACMIVIFLTIAYLKLLFLKKYS 655

## RESULT 13

ABP52127  
ID ABP52127 standard; protein; 655 AA.

XX AC ABP52127;

XX DT 10-OCT-2002 (first entry)

XX DE Homo sapiens ABC transporter ABCG2 protein SEQ ID NO:79.

XX DE ATP-binding cassette transporter; ABC transporter; modulation; D loop;  
 KW cancer; bacterial infection; fungal infection; protozoal infection;  
 KW antibacterial; fungicide; protozoaside.

XX OS Homo sapiens.

XX PN EP1217066-A1.

XX PD 26-JUN-2002.

XX PF 21-DEC-2000; 2000EP-00870316.

XX PR 21-DEC-2000; 2000EP-00870316.

XX PA (UYGE-) UNIV GENT.

XX DR WPI; 2002-550404/59.

XX PT Modulating activity of ATP-binding cassette (ABC) transporters by  
 PT influencing dimerization of nucleotide binding domains through use of D  
 PT loop sequence of an ABC transporter, or its antisense peptide or peptide  
 PT mimetic.

XX PS Disclosure; Fig 3; 290pp; English.

XX CC The present invention describes a method (M1) for modulating the activity  
 CC of ATP-binding cassette (ABC) transporters by influencing the  
 CC dimerisation of the nucleotide binding domains comprising the (a) a  
 CC polypeptide (polyP) consisting of 5-50 amino acids comprising the D loop  
 CC sequence of an ABC transporter (ABP52049 to ABP52091); (b) a polyP  
 CC consisting of the D loop sequence of an ABC transporter; (c) a peptide  
 CC mimetic or antisense peptide of (a) or (b). ABC transporters have  
 CC antibacterial, fungicide and protozoaside activities. (M1) is useful for  
 CC selectively modulating the activity of ABC transporters belonging to the  
 CC group of multidrug transporter/P-glycoproteins. Bacterial, fungal or  
 CC protozoal ABC transporters are involved in the infection of a mammal or  
 CC in the induction of resistance to antibiotics or drugs in a mammal. (M1)  
 CC is useful for preventing, treating or alleviating diseases associated  
 CC with functionality of an ABC transporter. ABP52092 to ABP52140 represent  
 CC ABC transporter proteins given in the exemplification of the present  
 CC invention

XX SQ Sequence 655 AA;

Query Match 99.4%; Score 3331; DB 5; Length 655;  
 Best Local Similarity 99.4%; Pred. No. 0;

	Matches	651;	Conservative	1;	Mismatches	3;	Indels	0;	Gaps	0;
QY	1	MSSSNVEVFI	PVSQNTNGFP	PATASNDL	KAFTEGAVLS	FHNICYRV	KLSGFL	PCRP	VE	60
DB	1	MSSSNVEVFI	PVSQNTNGFP	PATVSNDL	KAFTEGAVLS	FHNICYRV	KLSGFL	PCRP	VE	60
QY	61	KEILSNINGIM	KPGNAILG	PTGGKSSLL	DVLAARD	PSGLSG	DVLLNG	APRAN	FKCN	120
DB	61	KEILSNINGIM	KPGNAILG	PTGGKSSLL	DVLAARD	PSGLSG	DVLLNG	APRAN	FKCN	120
QY	121	SGYVVQDDV	VMGTLV	RENLOF	SAALRLAT	TMTNHEK	NERINRV	IQELGL	DKVAD	SKYT 180
DB	121	SGYVVQDDV	VMGTLV	RENLOF	SAALRLAT	TMTNHEK	NERINRV	IEELGL	DKVAD	SKYT 180
QY	181	QFIRVSGGER	KRTSIG	IMELITD	PSILF	DEPTTGL	SDSTAN	VLLIL	KRSKQ	RTIIF 240
DB	181	QFIRVSGGER	KRTSIG	IMELITD	PSILF	DEPTTGL	SDSTAN	VLLIL	KRSKQ	RTIIF 240
QY	241	SIHQPRYSI	FKLFD	SLTLL	ASGRML	FHGP	PAQ	EALGY	FESAG	VHCEAYNNPADFFLDIING 300
DB	241	SIHQPRYSI	FKLFD	SLTLL	ASGRML	FHGP	PAQ	EALGY	FESAG	VHCEAYNNPADFFLDIING 300
QY	301	DSTAVALN	REDFKATE	IIEPSK	QDKPLIE	KLAETI	YVNSS	FFYK	ETKAE	LHQLSGGKKKK 360
DB	301	DSTAVALN	REDFKATE	IIEPSK	QDKPLIE	KLAETI	YVNSS	FFYK	ETKAE	LHQLSGGKKKK 360
QY	361	ITVFEKEI	SVTTSC	HQLRW	SKSPK	NLLGN	PQASIA	QII	IVTVL	GLVIGAIYFGLKND 420
DB	361	ITVFEKEI	SVTTSC	HQLRW	SKSPK	NLLGN	PQASIA	QII	IVTVL	GLVIGAIYFGLKND 420
QY	421	TGIONRAG	VLPFLTT	NQCFSS	SAVELF	VVEKKL	FIHEYI	SGYRV	SSYF	LGLKLLSDLLP 480
DB	421	TGIONRAG	VLPFLTT	NQCFSS	SAVELF	VVEKKL	FIHEYI	SGYRV	SSYF	LGLKLLSDLLP 480
QY	481	MTMLPSI	IIFTCTI	VFMGLG	KPKAD	AFVMMFT	LMMVAYS	SASSMA	LAIAGQ	SVSVATLL 540
DB	481	MTMLPSI	IIFTCTI	VFMGLG	KPKAD	AFVMMFT	LMMVAYS	SASSMA	LAIAGQ	SVSVATLL 540
QY	541	MTICFVFM	IFSGLLV	NLTTIAS	LSWLQY	FSIPRY	GFTALQ	HNFFLQ	NFQCFGLNATGN 600	
DB	541	MTICFVFM	IFSGLLV	NLTTIAS	LSWLQY	FSIPRY	GFTALQ	HNFFLQ	NFQCFGLNATGN 600	
QY	601	NPCNYATCT	GSEYLV	KQIDLS	PLGMLK	QNHVAL	ACMIVIF	LTIAYL	KLLFLKKYS 655	
DB	601	NPCNYATCT	GSEYLV	KQIDLS	PLGMLK	QNHVAL	ACMIVIF	LTIAYL	KLLFLKKYS 655	

## RESULT 14

ABB07270  
ID ABB07270 standard; protein; 655 AA.

XX AC ABB07270;

XX DT 26-MAR-2002 (first entry)

XX DE Human BCRP (huhCRP) amino acid sequence.

XX KW Stem cell; ATP transport protein; ATP-binding cassette; antiparkinsonian;  
 KW hepatotropic; neurodegenerative; cytostatic; antianemic; muscular; BCRP;  
 KW cardiant; gene therapy.

XX OS Homo sapiens.

XX PN WO200192877-A2.

XX PD 06-DEC-2001.

XX PF 30-MAY-2001; 2001WO-US017459.

XX PR 31-MAY-2000; 2000US-00584586.

XX PR 29-MAY-2001; 2001US-00866866.

XX PA (SUJD-) ST JUDE CHILDREN'S RES HOSPITAL.

XX Sorrentino B, Schuetz J;  
XX WPI; 2002-114368/15.  
DR N-PSDB; ABA94369.  
XX Identifying a stem cell, for treating e.g., muscular dystrophy,  
PT myocardial infarction, Parkinson's disease, or neurodegenerative  
PT disorders, comprises detecting the expression of an ATP transport protein  
PT (BCRP) by a cell.  
XX Example; Page 78-80; 87pp; English.  
XX The invention provides a method of identifying and/or isolating a stem  
CC cell that involves detecting the expression of an ATP transport protein  
CC containing a conserved ATP-binding cassette (BCRP) by a cell in a sample  
CC comprising stem cells. The isolated stem cells may be used in the  
CC treatment of diseases such as muscular dystrophy, degenerative liver  
CC disorder, myocardial infarction, Parkinson's disease, degenerative  
CC disorders of the brain, and for tissue regeneration or replacement.  
CC Haematopoietic cells can be used in bone marrow transplants (e.g., for  
CC treatment of leukemia) and for ex vivo gene therapy for treating blood  
CC diseases such as sickle cell anemia and thalassemia. The stem cells can  
CC also be used as cell targets in gene therapy protocols. The present  
CC sequence represents the amino acid sequence of human BCRP  
XX Sequence 655 AA;  
SQ  
Query Match 99.4%; Score 3331; DB 5; Length 655;  
Best Local Similarity 99.4%; Pred. No. 0;  
Matches 651; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
QY 1 MSSSNVEVFIPVSGQNTGFPATASNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKVE 60  
Db 1 MSSSNVEVFIPVSGQNTGFPATVSNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKVE 60  
QY- 61 KEILSNINGIMKGLNAILGPTGGKSSLLDLVAARKDPSGLSGDVLINGAPRANFKCN 120  
Db 61 KEILSNINGIMKGLNAILGPTGGKSSLLDLVAARKDPSGLSGDVLINGAPRANFKCN 120  
QY 121 SGYVQDDVVMGTLTVRENIQFSAALRLATTTNHEKNERINRVIOELGLDKVADSKVGT 180  
Db 121 SGYVQDDVVMGTLTVRENIQFSAALRLATTTNHEKNERINRVIOELGLDKVADSKVGT 180  
QY 181 QFIRGVSGGRKRTSIGMELITDPSILFDEPTGLDSTANAVLLLLKMSKQGTIIF 240  
Db 181 QFIRGVSGGRKRTSIGMELITDPSILFDEPTGLDSTANAVLLLLKMSKQGTIIF 240  
QY 241 SIHQPRYSIFKLPDSLTLLASGRIMFHGPAQAEALGYFESAGYHCEAYNNPADFFLDIING 300  
Db 241 SIHQPRYSIFKLPDSLTLLASGRIMFHGPAQAEALGYFESAGYHCEAYNNPADFFLDIING 300  
QY 301 DSTAVALNRBEDFKATEIIBPSKQDKPLIEKLAETVYNSFYKETKAEHLQLSGGEKKK 360  
Db 301 DSTAVALNRBEDFKATEIIBPSKQDKPLIEKLAETVYNSFYKETKAEHLQLSGGEKKK 360  
QY 361 ITVPKEISYTTSPCHLRWYSKSFKNLLGNPQASIAQIIVTVVLGIVGAIYFGLKND 420  
Db 361 ITVPKEISYTTSPCHLRWYSKSFKNLLGNPQASIAQIIVTVVLGIVGAIYFGLKND 420  
QY 421 TGIQNRAGVLFFLTNQCFSVSVAVELFVVEKKLFIHEYISGYRVSSYFLGKLLSDLLP 480  
Db 421 TGIQNRAGVLFFLTNQCFSVSVAVELFVVEKKLFIHEYISGYRVSSYFLGKLLSDLLP 480  
QY 481 MTMLPSIIFTCIVVFMGLKPKADAFVVMFTLMMVAYSASSMALAIAGQSVSVATLL 540  
Db 481 MTMLPSIIFTCIVVFMGLKPKADAFVVMFTLMMVAYSASSMALAIAGQSVSVATLL 540  
QY 541 MTICFVFMIFSGLLVNLTTIASLWLSQYFSPRYGFTALQHNFEFLGQNCFCPLNATGN 600  
Db 541 MTICFVFMIFSGLLVNLTTIASLWLSQYFSPRYGFTALQHNFEFLGQNCFCPLNATGN 600  
QY 601 NPNYATCTGEEYLVKQIDLSLSPWGLMKNHVALACMIVIFLTIAYLKLLFLKKYS 655

Db 601 NPNYATCTGEEYLVKQIDLSLSPWGLMKNHVALACMIVIFLTIAYLKLLFLKKYS 655  
RESULT 15  
ABU63376  
ID ABU63376 standard; protein; 655 AA.  
XX AC ABU63376;  
XX 23-SEP-2003 (first entry)  
XX Human mitoxantrone resistance (MXR) transporter.  
XX Human; cytostatic; acridine derivative; MXR; Pgp; cancer;  
KW mitoxantrone resistance transporter; p-glycoprotein; lung cancer;  
KW colon cancer; breast cancer; prostate cancer; ovarian cancer;  
KW acute lymphocytic leukaemia; non-Hodgkin's lymphoma;  
KW multidrug resistance; MDR.  
XX Homo sapiens.  
XX US6521635-B1.  
XX 18-FEB-2003.  
XX 22-JAN-2001; 2001US-00767594.  
XX 20-JAN-2000; 2000US-0177410P.  
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.  
XX Bates S, Robey R;  
XX WPI; 2003-553619/52.  
XX New use of acridine derivative for inhibiting mitoxantrone resistance  
PT transporter in treatment of e.g. cancer, acute lymphocytic leukemia and  
PT non-Hodgkin's lymphoma.  
XX Disclosure; Col 29-34; 27pp; English.  
XX The invention relates to the use of an acridine derivative, or its salt  
CC or solvate, for inhibiting a mitoxantrone resistance (MXR) transporter in  
CC a cell over-expressing MXR gene but not over-expressing a p-glycoprotein  
CC (Pgp) gene, by contacting with the cell. Also included are assaying the  
CC modulation of the functional effect of a test compound on the cell by the  
CC acridine derivative (comprising contacting the test compound with the  
CC cells in the presence and absence of the acridine derivative and  
CC measuring the ability of the acridine derivative to modulate the  
CC functional effect of the test compound) and the treatment of cancer that  
CC over-expresses the (MXR) gene but not over-expresses the (Pgp) gene  
CC (comprising co-administering a chemotherapeutic recognized by the MXR  
CC transporter and the acridine derivative. The method is used for  
CC inhibiting the MXR transporter, and for treating cancer (e.g. lung  
CC cancer, colon cancer, breast cancer, prostate cancer, acute lymphocytic  
CC leukaemia, non-Hodgkin's lymphoma and ovarian cancer). The acridine  
CC derivative is a potent multispecific antagonist capable of inhibiting or  
CC reversing both Pgp-mediated and MXR-mediate multidrug resistance (MDR)  
CC phenotype. The present sequence is the human MXR transporter sequence  
XX Sequence 655 AA;  
SQ  
Query Match 99.4%; Score 3331; DB 6; Length 655;  
Best Local Similarity 99.4%; Pred. No. 0;  
Matches 651; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
QY 1 MSSSNVEVFIPVSGQNTGFPATASNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKVE 60  
Db 1 MSSSNVEVFIPVSGQNTGFPATVSNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKVE 60  
QY 61 KEILSNINGIMKGLNAILGPTGGKSSLLDLVAARKDPSGLSGDVLINGAPRANFKCN 120



Db 61 KEILSNINGIMKPGNAILGFTGGKSSLLDVLAAKDPGSLGSDVLINGAPRANFKCN 120  
Qy 121 SGYVQDDVNVGTLTVRENLOFSAAALRLATMTWHEKNERINRVIOELGLDKVADSKVGT 180  
Db 121 SGYVQDDVNVGTLTVRENLOFSAAALRLATMTWHEKNERINRVIOELGLDKVADSKVGT 180  
Qy 181 QPIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSSSTANAVLLLLKRMKSQGRITIF 240  
Db 181 QPIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSSSTANAVLLLLKRMKSQGRITIF 240  
Qy 241 SIHOPRYSIFKLPDSLTLLASGRMFHGPQALGYFESAGYHCEAYNNPADPFLDIING 300  
Db 241 SIHOPRYSIFKLPDSLTLLASGRMFHGPQALGYFESAGYHCEAYNNPADPFLDIING 300  
Qy 301 DSTAVALNREEDFKATEIIEPSKODKPLIEKLABIYVNSSFYKETAELHQLSGGEKKK 360  
Db 301 DSTAVALNREEDFKATEIIEPSKODKPLIEKLABIYVNSSFYKETAELHQLSGGEKKK 360  
Qy 361 ITVPKEISYTTSFCHQLRWKRSFKNLLGNPOASIAQIIVTVVLGLVIGAIYFGLKND 420  
Db 361 ITVPKEISYTTSFCHQLRWKRSFKNLLGNPOASIAQIIVTVVLGLVIGAIYFGLKND 420  
Qy 421 TGIQNRAGVLPFLTNQCFSSVSAVELFWVEKCLFIHEYISGYRVSSYFLGKLLSDLLP 480  
Db 421 TGIQNRAGVLPFLTNQCFSSVSAVELFWVEKCLFIHEYISGYRVSSYFLGKLLSDLLP 480  
Qy 481 MTMLPSIIIFTCIVYFMLGLPKADAFFVMMFTLMMVAYSASSMALAIAGOSVVSATLL 540  
Db 481 MTMLPSIIIFTCIVYFMLGLPKADAFFVMMFTLMMVAYSASSMALAIAGOSVVSATLL 540  
Qy 541 MTICFVFMWIFGGLVNLTTIASWLSWLYQYFSIPRYGFTALQHNEFLGQNCPCGLNATGN 600  
Db 541 MTICFVFMWIFGGLVNLTTIASWLSWLYQYFSIPRYGFTALQHNEFLGQNCPCGLNATGN 600  
Qy 601 NPCNYATCTGEEYLVKQIGIDLSFWGLWKNHVALACMIVIFLTIAVYKLLFLKKYS 655  
Db 601 NPCNYATCTGEEYLVKQIGIDLSFWGLWKNHVALACMIVIFLTIAVYKLLFLKKYS 655

Search completed: March 23, 2004, 18:29:34  
Job time : 64 secs